

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 10, 2004, 10:03:59 ; Search time 28 Seconds  
(without alignments)  
559.108 Million cell updates/sec

Title: US-10-081-644-2

Perfect score: 1920

Sequence: 1 MSVPTTQKAVIEGDKAVK.....ALTEGIEKGNKNVYARL 370

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 90 summaries

Database : Issued Patents AA.\*

- 1: /cgn2\_6/ptodata/2/iaa/5A COMB.pep.\*
- 2: /cgn2\_6/ptodata/2/iaa/5B COMB.pep.\*
- 3: /cgn2\_6/ptodata/2/iaa/6A COMB.pep.\*
- 4: /cgn2\_6/ptodata/2/iaa/6B COMB.pep.\*
- 5: /cgn2\_6/ptodata/2/iaa/PTCUS COMB.pep.\*
- 6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	266.5	13.9	363	4	US-09-215-694-5
2	202.5	10.5	340	4	US-09-107-532A-7298
3	173.5	9.0	350	4	US-09-328-352-8160
4	158.5	8.3	396	3	US-08-860-656B-2
5	158.5	8.3	396	3	US-08-860-656B-3
6	153.5	8.0	336	1	US-08-713-254-2
7	153.5	8.0	336	3	US-08-855-767-2
8	151.5	7.9	340	4	US-09-134-001C-3258
9	149.5	7.8	322	4	US-09-154-750A-81
10	141	7.3	349	4	US-09-107-532A-5838
11	139.5	7.3	344	4	US-09-252-991A-16756
12	134	7.0	401	4	US-09-252-991A-19274
13	133	6.9	6396	4	US-09-410-551B-92
14	131.5	6.8	396	3	US-08-860-656B-4
15	130.5	6.8	370	4	US-09-328-352-5285
16	126	6.6	2532	4	US-09-215-694-10
17	125.5	6.5	336	4	US-09-252-991A-11476
18	124.5	6.5	371	4	US-09-501-115-14
19	122.5	6.4	336	4	US-08-976-063E-20
20	121.5	6.3	412	4	US-09-252-991A-22667
21	121	6.3	345	2	US-07-857-224B-94
22	118	6.1	344	2	US-07-857-224B-93
23	115.5	6.0	2152	3	US-09-036-987A-3
24	115.5	6.0	2152	3	US-09-370-700-3
25	115.5	6.0	2152	4	US-09-603-207-3
26	114	5.9	3724	2	US-08-804-227C-10
27	114	5.9	3724	2	US-08-804-198-4

RESULT 1

US-09-215-694-5  
; Sequence 5, Application US/09215694B  
; Patent No. 6391583  
; GENERAL INFORMATION:

ALIGNMENTS

Sequence 5638, Ap  
Sequence 2, Appli  
Sequence 10, Appl  
Sequence 29389, A  
Sequence 92, Appl  
Sequence 95, Appl  
Sequence 32, Appl  
Sequence 4, Appl  
Sequence 8, Appl  
Sequence 3, Appl  
Sequence 7738, Ap  
Sequence 3374, Ap  
Sequence 216, App  
Sequence 216, App  
Sequence 2, Appl  
Sequence 2, Appl  
Sequence 14, Appl  
Sequence 28, Appl  
Sequence 96, Appl  
Sequence 24660, A  
Sequence 4988, Ap  
Sequence 2, Appl  
Sequence 32383, A  
Sequence 1085, Ap  
Sequence 2, Appl  
Sequence 2, Appl  
Sequence 7360, Ap  
Sequence 34, Appl  
Sequence 89, Appl  
Sequence 31262, A  
Sequence 2, Appl  
Sequence 33, Appl  
Sequence 3, Appl  
Sequence 2, Appl  
Sequence 6054, Ap  
Sequence 5, Appl  
Sequence 5, Appl  
Sequence 5, Appl  
Sequence 5, Appl  
Sequence 5, Appl  
Sequence 7, Appl  
Sequence 2, Appl  
Sequence 2, Appl  
Sequence 5343, Ap  
Sequence 12, Appl  
Sequence 4, Appl  
Sequence 4056, Ap  
Sequence 1, Appl  
Sequence 1, Appl  
Sequence 2, Appl  
Sequence 2, Appl  
Sequence 5503, Ap  
Sequence 7473, Ap  
Sequence 6, Appl  
Sequence 124, App  
Sequence 25683, A  
Sequence 4, Appl

4 US-09-107-532A-5638  
2 US-08-828-010-2  
2 US-09-501-115-10  
4 US-09-252-991A-29389  
4 US-07-857-224B-92  
2 US-07-857-224B-95  
4 US-09-501-115-32  
2 US-07-642-734C-4  
3 US-08-439-009A-4  
4 US-09-501-115-8  
3 US-09-428-517-3  
4 US-09-328-352-7738  
4 US-09-134-001C-3374  
1007 3 US-08-961-083-216  
4 US-09-536-784-216  
376 1 US-08-608-241-2  
376 2 US-08-922-182-2  
376 2 US-08-919-953-2  
376 3 US-09-192-983-2  
3562 4 US-09-679-279-14  
392 4 US-09-501-115-28  
346 2 US-07-857-224B-96  
448 4 US-09-328-352-4988  
358 4 US-08-805-191-2  
352 2 US-09-122-210-2  
460 4 US-09-198-452A-1085  
461 3 US-09-122-210-2  
461 3 US-09-443-681-2  
375 4 US-09-328-352-7360  
1095 4 US-08-851-567B-34  
324 4 US-09-154-750A-89  
351 4 US-09-252-991A-31262  
3739 3 US-09-320-878-2  
3739 3 US-09-105-537-33  
3739 4 US-09-141-908-3  
3739 4 US-09-657-440-2  
11877 3 US-09-105-537-6  
347 4 US-09-107-532A-6054  
7257 3 US-09-335-409-5  
7257 4 US-09-568-102-5  
7257 4 US-09-567-969-5  
7257 4 US-09-568-480-5  
7257 4 US-09-568-486-5  
7257 4 US-09-568-472-5  
7257 4 US-09-567-899-5  
1098 1 US-07-777-715-7  
1098 1 US-08-170-126-2  
1098 3 US-08-954-418-2  
344 4 US-09-134-001C-5343  
355 4 US-09-501-115-12  
3729 2 US-08-804-227C-4  
340 4 US-09-134-001C-4056  
2308 1 US-08-015-973-1  
2308 2 US-08-448-164-1  
2308 3 US-08-081-929-2  
2314 4 US-09-816-703A-2  
382 4 US-09-328-352-5503  
353 4 US-09-328-352-7473  
358 4 US-09-501-115-6  
520 4 US-09-724-623-124  
718 4 US-09-252-991A-25683  
360 4 US-09-501-115-4

28 113.5 5.9 200  
29 111 5.8 300  
30 110.5 5.8 366  
31 110.5 5.8 381  
32 108 5.6 344  
33 108 5.6 349  
34 108 5.6 360  
35 108 5.6 3567  
36 108 5.6 3567  
37 107.5 5.6 361  
38 107.5 5.6 3816  
39 106.5 5.5 382  
40 105.5 5.5 354  
41 105.5 5.5 1007  
42 105.5 5.5 1007  
43 105 5.5 376  
44 105 5.5 376  
45 105 5.5 376  
46 105 5.5 376  
47 104 5.4 3562  
48 102.5 5.3 292  
49 102.5 5.3 346  
50 102.5 5.3 448  
51 102 5.3 358  
52 101 5.3 352  
53 101 5.3 483  
54 100.5 5.2 460  
55 100.5 5.2 461  
56 100.5 5.2 461  
57 100 5.2 375  
58 100 5.2 1095  
59 99.5 5.2 324  
60 99 5.2 351  
61 98.5 5.1 3739  
62 98.5 5.1 3739  
63 98.5 5.1 3739  
64 98.5 5.1 3739  
65 98.5 5.1 11877  
66 97.5 5.1 347  
67 96.5 5.0 7257  
68 96.5 5.0 7257  
69 96.5 5.0 7257  
70 96.5 5.0 7257  
71 96.5 5.0 7257  
72 96.5 5.0 7257  
73 96.5 5.0 7257  
74 96 5.0 1098  
75 96 5.0 1098  
76 96 5.0 1098  
77 95.5 5.0 344  
78 95 4.9 355  
79 94.5 4.9 3729  
80 94 4.9 340  
81 94 4.9 2308  
82 94 4.9 2308  
83 94 4.9 2308  
84 94 4.9 2314  
85 93.5 4.9 382  
86 93 4.8 353  
87 93 4.8 358  
88 93 4.8 520  
89 93 4.8 718  
90 92.5 4.8 360



Query Match	9.0%;	Score 173.5;	DB 4;	Length 350;
Best Local Similarity	29.4%;	Pred. No. 3.4e-09;		
Matches	77;	Conservative	34;	Mismatches 106; Indels 45; Gaps 12;
QY	1	MSVPTTQKAVIIIEGDKAWKTDSVPPELKEGTALVKVEAVAGNPTDWKHIAKYIGPEG-	59	
Db	15	MKAVALQKAGPITLPEALVDIELDTTPVAKGHDLIVRVOAVSVPVDTK-IRKNVSAEQG	73	
QY	60	--ILGCDIAGTVVKVLGNASTDLKVGDTFGFVHGASQTDPKGAFAYARVYPPLFYKS	117	
Db	74	WKVLGMWDAVTVEAIGDKV-TQFKIGDV--VVYAGALN---RQSGNSBLQLV-----	119	
QY	118	NLTHSTADE--ISBGPKNFES--AASIPVSLTTAGVSLCHLGSKMWHSTPQHHTPL	173	
Db	120	-----DERIVGHKP-KTLTEATEAAALPLTAIWEMLFORLV-----PKTAPANTTI	166	
QY	174	LWGGATAVGQQLTQVAKHINAYTKIVTVASKHEKLKSYGADDDVFYHD--AGVIEQI	231	
Db	167	LVVIGGAGGVGSITQLLKQLTNLTIIATASPETKEWVEQLGADYVLDHRQPLAAQIKQL	226	
QY	232	KSKVP-----NLQHVIDAV	245	
Db	227	GLSAPLYVFSFTETDQHLSDIV	248	

```

RESULT 4
US-08-860-656B-2
; Sequence 2, Application US/08860656B
; Patent No. 6297028
; GENERAL INFORMATION:
; APPLICANT: Taniguchi, Tadatsugu
; APPLICANT: Shibuya, Hiroshi
; APPLICANT: Barsoumian, Edward L.
; TITLE OF INVENTION: IL-2R-Associated Polypeptides and DNA Molecules Coding Therefor
; FILE REFERENCE: 0652.1640000
; CURRENT APPLICATION NUMBER: US/08/860,656B
; CURRENT FILING DATE: 1998-01-15
; PRIOR APPLICATION NUMBER: PCT/EP95/05123
; PRIOR FILING DATE: 1995-12-23
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 396
; TYPE: PRT
; ORGANISM: Human p43
US-08-860-656B-2

```

Query Match	8.3%;	Score 158.5;	DB 3;	Length 396;
Best Local Similarity	27.8%;	Pred. No. 1.4e-07;		
Matches	71;	Conservative 39;	Mismatches 96;	Indels 49; Gaps 12
Qy	34	LVKVEAVAGNPTDMK-----HIAYKIGPEGSI-LGCDIAGTVVKLGPN	75	
Db	75	IVKVAASVNPIDVNRSGYCATALNMRDPLHWKIK-GEFFPLTLGRDVGSGVMECCLD	133	
Qy	76	ASTDLKVGDTGFGFVHGASQDTPKNGAFAYRVPPLFYKSNLTHSTADISEGPVK-N	134	
Db	134	VKY-FKPGDEW-----AAVPPWKQGLSEFVVV-----SGNEVSHKPKSLT	174	
Qy	135	PESAASIPLVSLTTAGVSLICHHLGSKMEWHPTPQHTHLLWGGATATVGCQQLIQVAKHN	194	
Db	175	HTQAASLPYVALTAW-SAINKVGGLNDKNCYTKR-----VLILGASGGVGTFAIQMKAMD	229	
Qy	195	AYTKIVTVASKKHEKLKLSGADDDVFDYHDAGVIEQIKSKYNLQHVLDVAGV--SEDSIP	252	
Db	230	AH--VTAVCSQDASELVKLGADDVIDYKSSVSEQLKSLXP-PDFILDNVGGSSTETWAP	286	
Qy	253	EAYKVTADSLPATLL	267	
Db	287	DFLKKWSGATVYTLV	301	

```

RESULT 5
US-08-860-656B-3
; Sequence 3, Application US/08860656B
; Patent No. 6297028
; GENERAL INFORMATION:
; APPLICANT: Taniguchi, Tadatsugu
; APPLICANT: Shibuya, Hiroshi
; APPLICANT: Barboomian, Edward L.
; TITLE OF INVENTION: 1b-2R-Associated Polypeptides and DNA Molecules Coding Therefor
; FILE REFERENCE: 0652.1640000
; CURRENT APPLICATION NUMBER: US/08/860,656B
; CURRENT FILING DATE: 1998-01-15
; PRIOR APPLICATION NUMBER: PCT/EP95/05123
; PRIOR FILING DATE: 1995-12-23
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 396
; TYPE: PRT
; ORGANISM: Human p43
US-08-860-656B-3

```

RESULT 6  
US-08-713-254-2  
; Sequence 2, Application US/08713254  
; Patent No. 5763236  
; GENERAL INFORMATION:  
; APPLICANT: KOJIMA, TOMOKO  
; APPLICANT: YAMAMOTO, HIROAKI  
; APPLICANT: KAWADA, NAOKI  
; APPLICANT: MATSUYAMA, AKINOSBU  
; TITLE OF INVENTION: NOVEL ENZYME, A METHOD TO PREPARE SAID  
; TITLE OF INVENTION: ENZYME, A DNA SEGMENT ENCODING SAID ENZYME, A TRANSFORMANT  
; TITLE OF INVENTION: CONTAINING SAID DNA SEGMENT AND A METHOD OF PREPARING  
; TITLE OF INVENTION: OPTICALLY ACTIVE ALCOHOL USING SAID ENZYME  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,  
; ADDRESS: P.C.  
; STREET: 1755 S. Jefferson Davis Highway, Suite 400  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22202  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/713,254

FILING DATE: 12-SEP-1996

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/311,328

FILING DATE: 23-SEP-1994

APPLICATION NUMBER: JP 5-261649

FILING DATE: 24-SEP-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 5-337191

FILING DATE: 28-DEC-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 6-181308

FILING DATE: 02-AUG-1994

ATTORNEY/AGENT INFORMATION:

NAME: Oblon, No. 5763236man F.

REGISTRATION NUMBER: 24,618

REFERENCE/DOCKET NUMBER: 6423-001-0

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 413-3000

TELEFAX: (703) 413-2220

TELEX: 248855 OPAT UR

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 336 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-713-254-2

Query Match 8.0%; Score 153.5; DB 1; Length 336;

Best Local Similarity 24.8%; Pred. No. 3.5e-07;

Matches 101; Conservative 50; Mismatches 125; Indels 131; Gaps 22;

Qy 1 MSVPTTQKAVIIIEGDKAV-VKTDVSVPELKEGTALVKVEAVAGNPTDWHKIAYK---IGP 56

Db 1 MSIPSSQYGFVFNKQSGNLNRNDLPVHKPKAGQLLLKVDVAGLCHSD-LHVIYEGDLCG- 58

Qy 57 EGSILGCDIAGTVVKLGPNASTDLKVGDT-----GFGFVHGASQTDPKN----- 100

Db 59 DNYVMGHEIAGTAAVAGDDV-INYKVGDRVACVGNCGCGCKYCRGAIDNVCKNAFGDWF 117

Qy 101 -----GAFAEYARVYPPFLFYKSNLTHSTADEISEGPKNFESAASLPVSLTTAGVSLCHH 155

Db 118 GLGYDGGYQQYLLVTRP-----RNLSR-IPDNVSAD-----VAAASTDAVLTP-----YH 161

Qy 156 LGSKMEWHPSTPQHTPLIIGWGATAVGQOLIQVAKHINAYTKIVTVASKKHE--KLLKS 213

Db 162 AIKMAQVSPT-----SNILLI--GAGGLGNAIQVAK--AFGAKVTVLDKKEARDQAKK 212

Qy 214 YGADDVFDYHDAGVIEQIKSKYPNLQHVDAVGSSEDSIPEAYKVTADSLPATLLEVVPM 273

Db 213 LGADAVY----- 219

Qy 274 IESIPPEIRKDNVKIDITLLYRASQEIILGATFRFPASPEYHE--ATVKFVKFINPHL-- 329

Db 220 -ETLPESISPGS-----FSACDFVSQVAT-FDVCQKYVPKGVMPVGLGAPNLSF 269

Qy 330 NNGDIHNMINKVF-----SNGLDDVPALTEGIEKGNKNKVKVVARL 370

Db 270 NLGLDALREIRILGSPFGTNDLDDVLKL---VSEGVKVPVRSACL 313

RESULT 7

US-08-855-767-2

Sequence 2, Application US/08855767

Patent No. 6255092

GENERAL INFORMATION:

APPLICANT: KOJIMA, TOMOKO

APPLICANT: YAMAMOTO, HIROAKI

APPLICANT: KAWADA, NAOKI

APPLICANT: MATSUYAMA, AKINOBU  
TITLE OF INVENTION: NOVEL ENZYME, A METHOD TO PREPARE SAID  
TITLE OF INVENTION: ENZYME, A DNA SEGMENT ENCODING SAID ENZYME, A TRANSFORMANT  
TITLE OF INVENTION: CONTAINING SAID DNA SEGMENT AND A METHOD OF PREPARING  
TITLE OF INVENTION: OPTICALLY ACTIVE ALCOHOL USING SAID ENZYME  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,  
ADDRESS: 1755 S. Jefferson Davis Highway, Suite 400  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/855,767

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/311,328

FILING DATE:

APPLICATION NUMBER: JP 5-337191

FILING DATE: 28-DEC-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 6-181308

FILING DATE: 02-AUG-1994

ATTORNEY/AGENT INFORMATION:

NAME: Oblon, No. 6255092man F.

REGISTRATION NUMBER: 24,618

REFERENCE/DOCKET NUMBER: 6423-001-0

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 413-3000

TELEFAX: (703) 413-2220

TELEX: 248855 OPAT UR

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 336 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-855-767-2

Query Match 8.0%; Score 153.5; DB 3; Length 336;

Best Local Similarity 24.8%; Pred. No. 3.5e-07;

Matches 101; Conservative 50; Mismatches 125; Indels 131; Gaps 22;

Qy 1 MSVPTTQKAVIIIEGDKAV-VKTDVSVPELKEGTALVKVEAVAGNPTDWHKIAYK---IGP 56

Db 1 MSIPSSQYGFVFNKQSGNLNRNDLPVHKPKAGQLLLKVDVAGLCHSD-LHVIYEGDLCG- 58

Qy 57 EGSILGCDIAGTVVKLGPNASTDLKVGDT-----GFGFVHGASQTDPKN----- 100

Db 59 DNYVMGHEIAGTAAVAGDDV-INYKVGDRVACVGNCGCGCKYCRGAIDNVCKNAFGDWF 117

Qy 101 -----GAFAEYARVYPPFLFYKSNLTHSTADEISEGPKNFESAASLPVSLTTAGVSLCHH 155

Db 118 GLGYDGGYQQYLLVTRP-----RNLSR-IPDNVSAD-----VAAASTDAVLTP-----YH 161

Qy 156 LGSKMEWHPSTPQHTPLIIGWGATAVGQOLIQVAKHINAYTKIVTVASKKHE--KLLKS 213

Db 162 AIKMAQVSPT-----SNILLI--GAGGLGNAIQVAK--AFGAKVTVLDKKEARDQAKK 212

Qy 214 YGADDVFDYHDAGVIEQIKSKYPNLQHVDAVGSSEDSIPEAYKVTADSLPATLLEVVPM 273

Db 213 LGADAVY----- 219

Qy 274 IESIPPEIRKDNVKIDITLLYRASQEIILGATFRFPASPEYHE--ATVKFVKFINPHL-- 329



```
Db 220 -ETLPESISPGS-----FSACFDVSVQAT-FDVCQKYVEPKGVIMPVGLGAENLSF 269
QY 330 NNGDIHHMIKVF-----SNGLDDVPALTEGKEGKNKVKVAVRL 370
Db 270 NLGDLALREIRILGSPWGTNDLDDVLKL----VSEKVKPVVRSACL 313

RESULT 8
US-09-134-001C-3258
; Sequence 3258, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3258
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3258

Query Match
Best Local Similarity 7.9%; Score 151.5; DB 4; Length 340;
Matches 67; Conservative 33; Mismatches 102; Indels 39; Gaps 11;

QY 6 TQKAVLIEGKAVKTD-----VSPDELKEGTALVKVEAVAGNPTDKHIAIKGP 56
Db 4 TWKAI---GFKSSFQDDEGNCFEFNFDIPHPSGHELLVKVQSVNPDVKQRTMPVDK 60
QY 57 EGSILGCDIAGTVVKLGPNASTDOLKVDGTGFGFVHGASOTDPKNGAFAYARVVPPLFYK 116
Db 61 APRVLGFDVAVGIEKIGDQVSM-FQEGDVVF-----YSGSPNQNGSNEEVLIEEYL--- 111
QY 117 SNLTHSTADEISSEGPVK-NFESAASLPVSLTTAGVSLCHHLGSKMEWHPTPQHTHPLLI 175
Db 112 -----VAKAPTNLSEQAASLPGLTGLTAVETLFDVFGISKE--PS-ENKGSLLI 158
QY 176 WGGATAVGQOLIQVAKHINATKIYTVASKKHE-KLLKSYGADDVFDYHDAGVIEQIKSK 234
Db 159 INGAGGVGIATQIARFYG--LKVITTSAREDTIKWSVNNMGADVVLN-HKKDLISQOFKON 215
QY 235 Y 235
Db 216 H 216

RESULT 9
US-09-154-750A-81
; Sequence 81, Application US/09154750A
; Patent No. 6432640
; GENERAL INFORMATION:
; APPLICANT: Vogelstein, Bert
; APPLICANT: Kinzler, Kenneth
; APPLICANT: Polyak, Kornelia
; TITLE OF INVENTION: p53-Induced Apoptosis
; FILE REFERENCE: 1107.75357
; CURRENT APPLICATION NUMBER: US/09/154,750A
; CURRENT FILING DATE: 1998-09-17
; PRIOR FILING DATE: 1998-09-17
; PRIOR APPLICATION NUMBER: 60/059,153
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/079817
; PRIOR FILING DATE: 1998-03-30
; NUMBER OF SEQ ID NOS: 93
; SOFTWARE: FastSeq for Windows Version 3.0
```

```
; SEQ ID NO 81
; LENGTH: 322
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-154-750A-81

Query Match
Best Local Similarity 7.8%; Score 149.5; DB 4; Length 322;
Matches 64; Conservative 34; Mismatches 105; Indels 35; Gaps 9;

QY 14 GDKAVVKTDSVPELKEGTALVKVEAVAGNPTDKHIAIKGP---EGSILGCDIAGTVV 70
Db 11 GPNLYVKEVAKSPGEGEVLLKVAASALNRADLMORQGOYDPPPGASNLGLEASGHA 70
QY 71 KLGPNASTDOLKVDGTGFGFVHGASOTDPKNGAFAYARVVPPLFYKSNLTHSTADEISEG 130
Db 71 ELGPGCGQHWKIGDTAMALLPGGQ-----AQYTVPEGLM-----PIPEG 112
QY 131 PVKNFESAASLPVSLTTAGVSLCHHLGSKMEWHPTPQHTHPLLIWGGATAVGQOLIQVA 190
Db 113 --LTLTQAAAIPAEAWLTA-FQLLHLVGN-----VQAGDYVLIHAGLSGVGTAAIQLT 161
QY 191 KHINATKIYTVASKKHEKLLKSYGADDVFDYHDAGVIE-QIK-SKYPNLOHVDAVG 246
Db 162 RMAGA-IPLVTAGSQKQLQMAEKLGAAGFNKYKEDFSEATLKTGAGVNLILDCIG 218

RESULT 10
US-09-107-532A-5838
; Sequence 5838, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; CORRESPONDENCE ADDRESS: ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; ADDRESS: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 5838:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 349 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
```

```
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...349
; SEQUENCE DESCRIPTION: SEQ ID NO: 5838:
US-09-107-532A-5838

Query Match      7.3%; Score 141; DB 4; Length 349;
Best Local Similarity 22.5%; Pred. No. 6.9e-06;
Matches 89; Conservative 52; Mismatches 148; Indels 106; Gaps 17;

QY 1 MSVPTTKAVILIGDKRAVKTDSVPELKEGTALVKVEAVAGNPTDKHIAIKIGEGS- 59
DB 11 MKVVGPFYEGLPIDDPKSFIDEKQIPVPSRDLLVKVNAVSNVPTK-LRQDNGIRNAL 69
QY 60 -ILGCDIAGTVKLGPNASTDLKVGDTGFGFVHGASQTDPKNGAFAYEYARVYPPLEKSN 118
DB 70 RILGFGVGKVAAGVEQV-KFSVGDRVP-----YAGTTTRAGSNQEQYQLV----- 114
QY 119 LTHSTADE--ISRGPKVNF--ESAASLPVSLTTAGVSLCHHLGSKMEWHPSTPQHT- 170
DB 115 -----DERIVALAP-KNLSDEEAALPLTSLTAYELLPEKFG-----LTPEENANRG 160
QY 171 HPLLWGGATAVGOQLIQVAKHINAYTKIVTVASKKHEKLLKSYGADDVDFYHD- 224
DB 161 KKLIVNGSGVGSILNQLAHW--AGLEVYATASPKNFELKTKTGVDYDIDYHDLKSGF 218
QY 225 -----AGVIEQIKSKYVNLQHVDAVGSSEDSIPEAYKVTADSLPATLLEVVPMTI 274
DB 219 RELGIDQVEVAVLFDITRFDOIKHLRPFPHVG-----AKTDYVETIQTGEALAHIA 300
QY 275 ESIPERI-----RDNVKIDITILLYRASGOEILLGATRRPPASPEYH---EATVKFVKFIN 326
DB 256 VGIEKPLDIDGWNKQISPDWEYMF-----AKTDYVETIQTGEALAHIA 300
QY 327 PHLWNGDIHNMNIVKPSNGLD--DVPALTEGIKEG 359
DB 301 ALANEGIIIRSTLKVSYSDGNKAKNLKQATKDVETG 335

RESULT 11
US-09-252-991A-16756
; Sequence 16756, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 16756
; LENGTH: 344
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16756

Query Match      7.0%; Score 134; DB 4; Length 401;
Best Local Similarity 21.3%; Pred. No. 4.5e-05;
Matches 73; Conservative 54; Mismatches 150; Indels 66; Gaps 13;

QY 14 GDKAVKTDVSVPELKEGTALVKVEAVAGNPTDKHIAIKIG-----PEGS--ILGCDIAG 67
DB 84 GPEVLVATSRPLTPGPREVLVEVRAAGVNGPD---VLQKGVYDPPPGASDIPGLEIAG 140
QY 68 TVVKLGPNASTDLKVGDTGFGFVHGASQTDPKNGAFAYEYARVYPPLEKSNLTHSTADE- 126
DB 141 MYRAGSEVSR-FAVGEAVMALIPG-----GGYAQFA-----VADER 176
QY 127 ----ISEGPKVNFESAASLPVSLTTAGVSLCHHLGSKMEWHPSTPQHTPLLIWGGATAV 182
DB 177 TTLHLPDG--LCWEERAALPEFTMTVWNLFORGGFKA-----GETLLVHGASGI 225
QY 183 GOOLIQVAKHINAYTKIVTVASKKHEKLLKSYGADDVDFYHDAGVIEQI--KSKYPNLQH 240
DB 226 GTAATMLGKAFGAAKIFTTISSEAQREASRLGADLAINYTEQDFVEEVLRTGREGVDV 285
QY 241 VIDAVGSEDSIPEAYKVTADSLPATLLEVVPMTIISIPEIRKDNVKIDITILLYRASGOE 300
DB 286 IVDIVAG-DYVTRNYQAAAMNGRIVOIGVI-----KGKAAEVDLFFPMLSKRL 331
QY 301 ILLGAT-RPPASPEYHEATVKVEKFINPHLNNGDIHNMNIVK 342
DB 332 VHLGSTLSRSRSHDEKAGIITAELEQWPHVRAGAVRPQVFRTP 374

RESULT 13
US-09-410-551B-72
; Sequence 72, Application US/09410551B
; Patent No. 6503737
; GENERAL INFORMATION:
; APPLICANT: KOSAN BIOSCIENCES, Inc.
; APPLICANT: REEVES, CHRISTOPHER
; APPLICANT: CHU, DANIEL
; APPLICANT: KHOSLA, CHAITAN
; APPLICANT: SANTI, DANIEL
; APPLICANT: WU, KAI
; TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA
; TITLE OF INVENTION: CONSTRUCTS THEREFOR
; FILE REFERENCE: 30062-20026.00
```



; SEQ ID NO 10  
; LENGTH: 2532  
; TYPE: PRT  
; ORGANISM: Aspergillus terreus  
US-09-215-694-10

Query Match 6.6%; Score 126; DB 4; Length 2532;  
Best Local Similarity 23.4%; Pred. No. 0.0056;  
Matches 56; Conservative 39; Mismatches 92; Indels 52; Gaps 10;

Qy 18 VVKTDVPELKEGTALVK--VEAVAGN--PTDW-----KHIAYKIGP-EGSI 60  
Db 1816 LLRLDIQTPGLDLSLHFTKRNVDYEPDKLPDDWVEIEPRAFGNLFRIIMVAMQLESNV 1875  
Qy 61 LGCDIAGTVVKLGNAST--DLKVGDTGFGFVHGASQTPKNGAFAYARVPPLEYS 117  
Db 1876 MGFECAVVTLSLSTARTIAPGLAVGDRVCALMNG-----HWAS 1914  
Qy 118 NLTHSTADEISEGPKVKNFESASLPVSLTTAGVSLCHHLGSKMEWHPSTPQHTPLIWMG 177  
Db 1915 RVTTSRNVVRIPETLSFPFHAASIPLAFTTAYISL--YTARI-----LPGET--VLIHA 1965  
Qy 178 GATAVGQOLIQAQKHINAYTKIVTVASKHEKLKSY--GADVDFYHDAGVIEQIKSK 234  
Db 1966 GAGVGQAAILAQLTGAEVFTTAGSETKRNLLIDKFLDPDHFVSSRDSFVDGIKTR 2024

RESULT 17  
US-09-252-991A-31476  
; Sequence 31476, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252.991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 31476  
; LENGTH: 336  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-31476

Query Match 6.5%; Score 125.5; DB 4; Length 336;  
Best Local Similarity 24.4%; Pred. No. 0.00025;  
Matches 76; Conservative 41; Mismatches 99; Indels 95; Gaps 17;

Qy 31 GTALVKVEAVAGNPTDWKHIAYKIG-----PEGSILGCDIAGTVVKLGNASTDLKVGDTG 86  
Db 41 GEVWLEQAAGVNFELD---LQORSQAVPIPLPSGLGLEGAGVVAALGPGVS--GLAPCD-- 94  
Qy 87 FGFVHGASQTPKNGAFAYARVPPLEPYKSNLTHSTADEISEGP--VKNFESASLPVSL 145  
Db 95 ----RVAYATGPL--GAYAS--ARLYP-----AERLLKLPDTLAFEDAAALLFKG 136  
Qy 146 TTAGVSLCHHLGSKMEWHPSTPQHTP-----LLIWWGATVCGQOLIQAQKHINAYTKI 199  
Db 137 ITA-----HYL-----LYATYPVPGTGRILLYGAAGAVGQQLMAAWARHLGAW--V 179  
Qy 200 TVVASKKHE--KLKSYGADVDFYHDAGVIEQI-----KX 233  
Db 180 IGUVSKAESVERAAGCDEVLFVDAASLAQAQVAELTAGRKVDVVDYDPIGRATFEASLNS 239  
Qy 234 KYPNLQHVDAVSEDSIPKAYKVTADSLPATLLEVVPMTI--ESPIEIRKONVKIDIT 291  
Db 240 LRP--RGLLVFGATGVPVPAVEATLNAGSLFLTRPSLAHAHTANPEE-----286

Query Match 6.5%; Score 124.5; DB 4; Length 371;  
Best Local Similarity 22.0%; Pred. No. 0.00036;  
Matches 74; Conservative 54; Mismatches 149; Indels 59; Gaps 15;

Qy 46 DMKHIAYKIGPEGSILGCDIAGTVVKLGNASTDLKVGDTGFG--FVHGASQTPDKNGAF 103  
Db 65 DMKNAYKIMP-----GHEIAGEVTEVGRNV--TKFAGDRVGVGCMVNSQSCSCDKGF 118  
Qy 104 AEYARVYPPLEFYKSN-----LTH---STADEISEGPKVKNFESASLP--VSLTTAGVS 151  
Db 119 ENHC---PGMIFTYNSVDRDGTTHGGYSSVMVWHFVVRFPDAMPLOKGAPELLCAGIT 175  
Qy 152 LCHHLGSKMEWHPSTPQHTPLIWWGATVCGQOLIQAQKHINAYTKIVTVASKHEKLK 211  
Db 176 VY-----SPMKYTHGLNAPGNHGLV--GLGGLGHVAVKFGAFKGMKVTVISSPGKQEAL 229  
Qy 212 KSYGADDDVFDHAGVIEQIKSKYPNLQHVDAVSEDSIPEA-----YKVTADSL 262  
Db 230 ERIGADAFVVSQDA---DEMKAATMTMDGIINTVSA--NVPWAPLEGLLKPNGKIMVGL 284  
Qy 263 PATLLEVVPMTIESIPEEIRKONVKIDITLLYRAG-----OEILIGATRFPPASPEYHEA 317  
Db 285 PEKPIEVPPFAL-----VARNKTLAGSCIGGMRDTQEMLDLAAKHGVTADIEVI 333  
Qy 318 TVKVFVFINPHLNGDIHMHNIKVFNSGLDDVPALT 353  
Db 334 GAERYNTAMERLAKADVRYRVIDIANTLTKAAAAT 369

RESULT 19  
US-08-976-063E-20  
; Sequence 20, Application US/08976063E  
; Patent No. 6524831  
; GENERAL INFORMATION:  
; APPLICANT: Steinbuechel, Alexander  
; APPLICANT: Priefert, Horst  
; APPLICANT: Rabenhorst, Jurgen  
; TITLE OF INVENTION: SYNTHETIC ENZYMES FOR THE PRODUCTION OF CONIFERYL  
; TITLE OF INVENTION: ALCOHOL, CONIFERYLALDEHYDE, FERULIC ACID, VANILLIN AND  
; TITLE OF INVENTION: VANILLIC ACID AND THEIR USE  
; FILE REFERENCE: Bayer-9998-CAO  
; CURRENT APPLICATION NUMBER: US/08/976.063E  
; CURRENT FILING DATE: 1997-11-21  
; PRIOR APPLICATION NUMBER: 196 49 655.1 GERMANY  
; PRIOR FILING DATE: 1996-11-29  
; NUMBER OF SEQ ID NOS: 45



Qy	54	I	----	CPGSIILCGD	IAGTVV	KLGPNAST	DLKVGD	-TGFGFVHCASOT	-----	96																																															
Db	61	L	P	L	V	G	H	E	G	-----																																															
Qy	97	-----	DPKNGAF	AEYAR	VY	P	PFYK	SNLTH	STAD	EISEG	PVKNF	ESAASL	PSVL	STT	AG	149																																									
Db	113	P	D	A	L	S	G	Y	T	H	D	S	F	Q	F	-----																																									
Qy	150	V	S	L	C	H	H	L	G	-----	SKMEW	P	S	T	P	Q	HTH	P	L	L	I	W	G	G	A	T	A	N	G	Q	L	I	Q	V	A	K	H	I	N	A	T	K	I	V	T	V	A	S	K	205							
Db	157	V	T	V	Y	K	A	L	K	B	A	D	L	K	A	G	D	W	-----	V	A	I	S	G	A	A	G	G	L	S	A	V	O	T	A	M	Y	R	V	L	G	I	D	A	G	E	206										
Qy	206	K	H	E	K	L	K	S	G	A	D	D	V	F	D	I	H	D	A	G	V	I	E	Q	I	K	S	Y	N	L	O	R	V	D	A	V	G	S	E	D	S	I	P	E	A	Y	K	V	T	A	D	S	L	P	A	T	265
Db	207	E	K	E	K	-----	K	L	G	G	V	F	I	D	F	-----	I	K	T	K	N	W	-----	V	S	D	I	E	A	T	K	-----	G	P	H	G	241																				
Qy	266	L	L	E	V	P	M	T	I	E	S	I	P	E	I	R	K	D	N	V	K	I	D	I	T	L	L	-----	R	A	S	G	E	I	L	L	G	A	T	P	P	A	S	-----	P	E	Y	H	E	A	T	K	320				
Db	242	V	I	N	V	-----	S	V	S	E	-----	A	A	I	S	T	S	I	V	R	P	C	G	T	V	V	W	-----	L	P	A	N	A	Y	K	V	E	S	F	V	H	V	K	286													
Qy	321	F	V	K	F	I	N	P	H	L	N	G	D	I	H	M	N	I	K	V	F	S	N	G	L	D	D	V	P	A	L	T	E	G	I	K	E	358																			
Db	287	S	I	N	I	G	S	V	G	N	A	D	R	E	A	L	D	F	S	R	G	L	I	K	S	P	I	K	I	V	G	L	S	E	324																						

RESULT 22  
US-07-857-224B-93  
; Sequence 93, Application US/07857224B  
; Patent No. 5958784  
; GENERAL INFORMATION:  
; APPLICANT: Benner, Steven A.  
; TITLE OF INVENTION: Predicting Folded Structures of Proteins  
; NUMBER OF SEQUENCES: 114  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Steven A. Benner  
; STREET: Hadlaubstrasse 151  
; CITY: Zurich  
; STATE: none  
; COUNTRY: Switzerland  
; ZIP: (note: this is an international post code) CH-8092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 inch diskette, 1.4 Mb storage  
; COMPUTER: Apple Macintosh  
; OPERATING SYSTEM: Macintosh 7.0  
; SOFTWARE: Microsoft Word  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/857,224B  
; FILING DATE: 03/25/92  
; CLASSIFICATION: 436  
; PRIOR APPLICATION DATA: none  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (International) 41 1 632 2830  
; TELEFAX: (International) 41 1 262 2437  
; TELEX: none  
; INFORMATION FOR SEQ ID NO: 93:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 344  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE:  
; DESCRIPTION: protein  
; ORIGINAL SOURCE:  
; ORGANISM: Aspergillus  
; FEATURE: Alcohol dehydrogenase, Table 3 Column 16  
; PUBLICATION INFORMATION:  
; AUTHORS:  
; AUTHORS: Joernvall, H.  
; AUTHORS: Persson, M.  
; AUTHORS: Jeffery, J  
; TITLE: Alcohol dehydrogenases  
; JOURNAL: Proceedings of the National Academy of Sciences, 1  
; VOLUME: 78  
; PAGES: 4226-4230

; DATE: 1981  
US-07-857-224B-93

Query Match      6.1%; Score 118; DB 2; Length 344;  
Best Local Similarity 21.0%; Pred.No. 0.0015;  
Matches 86; Conservative 52; Mismatches 147; Indels 124; Gaps 19;

Qy 2 SVPTTKAVII-BGDKAVVKTDVSPPELKEGTALVAVEA VAGNPT-----DWKHIAYK 53  
| : | : | : | : | : | : | : | : | : | : | : | : | : |  
Db 1 SIPETKAIIIFESNGKLKHIDIPVPKPENELLINVKYSGVCHTDLHAWHGDP-LPTK 59  
| : | : | : | : | : | : | : | : | : | : | : | : | : |  
  
Qy 54 I-----GPEGILCDIAGTVVKLGPNASTDLKVGD-TGGFVHGASQT-----96  
| : | : | : | : | : | : | : | : | : | : | : | : | : |  
Db 60 LPLVGHEG-----AGVVVGGENVK-GWIKIGDYAGIKWLNSCMACEYCGLGNESC 111  
| : | : | : | : | : | : | : | : | : | : | : | : | : |  
  
Qy 97 -----DPKNGAFAEYARVYPFLFYKSNIETHSTADEISGPVNPFESAASLPVSLLTAG 149  
| : | : | : | : | : | : | : | : | : | : | : | : | : |  
Db 112 PHADLSGYTHDGFSQEYA-----TADA-----QAHHIQGTDLAE 147  
| : | : | : | : | : | : | : | : | : | : | : | : | : |  
  
Qy 150 VS--LCHELLSKMEHPSTPQHPTHPLLIMGATAVGOOLTIQAKHNATKI-VTVASKK 206  
| : | : | : | : | : | : | : | : | : | : | : | : | : |  
Db 148 VAPILCAGITVYKALKSANLRAGHWAAISGAAGGLGS LAVOYAKAMGYRVLGIDGGPGK 207  
| : | : | : | : | : | : | : | : | : | : | : | : | : |  
  
Qy 207 HEKKLSYGADDVFVDHXDVIEQIKSYNPLOHVTDVGSSESIPEAYKYVTADS LPATL 266  
| : | : | : | : | : | : | : | : | : | : | : | : | : |  
Db 208 EETSLE---GGEVF-----IDFTKEKDIVSAVVKATNGGAHGII 242  
| : | : | : | : | : | : | : | : | : | : | : | : | : |  
  
Qy 267 -LEVWPMTIESIEIRKONVKIDITLLYRASSOEILLGATRPA----SPEYHEATVKF 321  
| : | : | : | : | : | : | : | : | : | : | : | : | : |  
Db 243 NVSVSEAIAEA-----STRYCRANGTVVVVG---LPAGAKCSSDFNVNHVKS 286  
| : | : | : | : | : | : | : | : | : | : | : | : | : |  
  
Qy 322 VKFINPHLNMGDIHHNNIKVFSN-----GLDDVPALTGIKEGK 360  
| : | : | : | : | : | : | : | : | : | : | : | : | : |  
Db 287 ISIVGSVGNRADTRREALOFFARGLVKSPIKVGLSSLPEIVEKMEKGQ 335  
| : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 23  
US-09-036-987A-3  
Sequence 3, Application US/09036987A  
Patent No. 6143526  
GENERAL INFORMATION:  
APPLICANT: Baltz, Richard H.  
APPLICANT: Broughton, Mary C.  
APPLICANT: Crawford, Kathryn P.  
APPLICANT: Madduri, Krishnamurthy  
APPLICANT: Merlo, Donald J.  
APPLICANT: Treadway, Patti J.  
APPLICANT: Turner, Jan R.  
APPLICANT: Waldron, Clive  
TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide  
NUMBER OF SEQUENCES: 39  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dow AgroSciences LLC Patent Department  
STREET: 9330 Zionsville Road  
CITY: Indianapolis  
STATE: Indiana  
COUNTRY: USA  
ZIP: 46268  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/036,987A  
FILING DATE: 09-MAR-1998  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Stuart, Donald R  
REGISTRATION NUMBER: 28,479  
REFERENCE/DOCKET NUMBER: 50,608  
TELECOMMUNICATION INFORMATION:

```

; TELEPHONE: (317)337-4816
; TELEFAX: (317)337-4847
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2152 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-036-987A-3

Query Match      6.0%; Score 115.5; DB 3; Length 2152;
Best Local Similarity 22.6%; Pred. No. 0.05;
Matches 79; Conservative 48; Mismatches 129; Indels 93; Gaps 19;

QY 17 AVKTDVSVPELKEGTALVKVEAVAGNPTDKHIAKYKIGPEGSLGCDIAGTVVKLGPN 76
Db 1419 ALVDEPTATAPLGDGEVRIAMRAAGVNFDRD-ALIALGMYPGVASLGSEGAGVVETGPGV 1477

QY 77 STDLKVGDTGFGFVHGASQTDPKNGAFAPAEYARVYPLFYKSNLTHSTADBISEGPKVNF 136
Db 1478 -TGLAPGDRVMGMI-----PK-----AFGPL-----AVADHRMVTTRIPAG--WSFA 1515

QY 137 SAASLPVSLTTAGVSLCHHLGSKMWHPTPQHTHPLLINGGATAVGQOLIQAKHINAY 196
Db 1516 RAASVPIVFLTAYALVDLAGLR-----PG-----ESLLVHSAAGGVGMAAIQIARHLGA- 1565

QY 197 TKIVTVASK-----KHEKLLKS-----YGADDVDF-----DYHDAGV- 227
Db 1566 -EYATASEDKQWAVELSHREHLASSRTCDPEQOFLGATGGRGVDDVVLNSLAGFADASLR 1624

QY 228 -----IE-----QIKSKYPNLOH-VIDAV-GSEDSIPPEAYKVYKVTADSLPATLL 267
Db 1625 MLPRGGRFLELGTVDVRDPEVADAHPGVSYQAFDTVEAGPQRIGEMLHVELFEGRVL 1684

QY 268 EVVPMT---IESIPPEIR-----KONVKIDITL--LYRASGOBILIGAT 306
Db 1685 EPLPVTAMDVRQAPALRHLSQLRHSQARHVGKULVTMPVWDAAAGTVLVTGGT 1733

RESULT 25
US-09-603-207-3
; Sequence 3, Application US/09603207B
; Patent No. 6521406
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H
; APPLICANT: Broughton, Mary C
; APPLICANT: Crawford, Kathryn P
; APPLICANT: Madduri, Krishnamurthy
; APPLICANT: Treadway, Patti J
; APPLICANT: Turner, Jan R
; APPLICANT: Waldron, Clive
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
; FILE REFERENCE: 50489 DIV1
; CURRENT APPLICATION NUMBER: US/09/603,207B
; CURRENT FILING DATE: 2000-06-23
; EARLIER APPLICATION NUMBER: 09/370,700
; EARLIER FILING DATE: 1998-03-09
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 2152
; TYPE: PRT
; ORGANISM: Saccharopolyspora spinosa
; US-09-603-207-3

Query Match      6.0%; Score 115.5; DB 4; Length 2152;
Best Local Similarity 22.6%; Pred. No. 0.05;
Matches 79; Conservative 48; Mismatches 129; Indels 93; Gaps 19;

QY 17 AVKTDVSVPELKEGTALVKVEAVAGNPTDKHIAKYKIGPEGSLGCDIAGTVVKLGPN 76
Db 1419 ALVDEPTATAPLGDGEVRIAMRAAGVNFDRD-ALIALGMYPGVASLGSEGAGVVETGPGV 1477

QY 77 STDLKVGDTGFGFVHGASQTDPKNGAFAPAEYARVYPLFYKSNLTHSTADBISEGPKVNF 136
Db 1478 -TGLAPGDRVMGMI-----PK-----AFGPL-----AVADHRMVTTRIPAG--WSFA 1515

QY 137 SAASLPVSLTTAGVSLCHHLGSKMWHPTPQHTHPLLINGGATAVGQOLIQAKHINAY 196
Db 1516 RAASVPIVFLTAYALVDLAGLR-----PG-----ESLLVHSAAGGVGMAAIQIARHLGA- 1565

QY 197 TKIVTVASK-----KHEKLLKS-----YGADDVDF-----DYHDAGV- 227
Db 1566 -EYATASEDKQWAVELSHREHLASSRTCDPEQOFLGATGGRGVDDVVLNSLAGFADASLR 1624

QY 228 -----IE-----QIKSKYPNLOH-VIDAV-GSEDSIPPEAYKVYKVTADSLPATLL 267
Db 1625 MLPRGGRFLELGTVDVRDPEVADAHPGVSYQAFDTVEAGPQRIGEMLHVELFEGRVL 1684

QY 268 EVVPMT---IESIPPEIR-----KONVKIDITL--LYRASGOBILIGAT 306
Db 1685 EPLPVTAMDVRQAPALRHLSQLRHSQARHVGKULVTMPVWDAAAGTVLVTGGT 1733

RESULT 24
US-09-370-700-3
; Sequence 3, Application US/09370700
; Patent No. 6274350
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H
; APPLICANT: Broughton, Mary C
; APPLICANT: Crawford, Kathryn P
; APPLICANT: Madduri, Krishnamurthy
; APPLICANT: Treadway, Patti J
; APPLICANT: Turner, Jan R
; APPLICANT: Waldron, Clive
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
; FILE REFERENCE: 50489 DIV1
; CURRENT APPLICATION NUMBER: US/09/370,700
; CURRENT FILING DATE: 1999-08-09
; EARLIER APPLICATION NUMBER: US 09/36987
; EARLIER FILING DATE: 1998-03-09
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 2152
; TYPE: PRT
; ORGANISM: Saccharopolyspora spinosa
; US-09-370-700-3

Query Match      6.0%; Score 115.5; DB 3; Length 2152;
Best Local Similarity 22.6%; Pred. No. 0.05;
Matches 79; Conservative 48; Mismatches 129; Indels 93; Gaps 19;

QY 17 AVKTDVSVPELKEGTALVKVEAVAGNPTDKHIAKYKIGPEGSLGCDIAGTVVKLGPN 76
Db 1419 ALVDEPTATAPLGDGEVRIAMRAAGVNFDRD-ALIALGMYPGVASLGSEGAGVVETGPGV 1477
```

Db 1685 EPLPVTAMDVRQAPALRHLHQARHVGKLVLTMPVWDAAGTGLVLTGGT 1733

## RESULT 26

US-08-804-227C-10  
; Sequence 10, Application US/08804227C  
; Patent No. 5876991  
; GENERAL INFORMATION:  
; APPLICANT: DeHoff, Bradley S.  
; APPLICANT: Kuhstoss, Stuart A.  
; APPLICANT: Rostock, Paul R., Jr.  
; APPLICANT: Sutton, Kimberly L.  
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: THOMAS G. PLANT 1501  
; STREET: LILLY CORPORATE CENTER  
; CITY: INDIANAPOLIS  
; STATE: IN  
; COUNTRY: USA  
; ZIP: 46285  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: ASCII(DOS) Text only  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/804,227C  
; FILING DATE: February 21, 1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Plant, Thomas G.  
; REGISTRATION NUMBER: 35,784  
; REFERENCE/DOCKET NUMBER: X-8231  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 317-276-2459  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3724 amino acids  
; TYPE: amino acid  
; TOPOLOGY: unknown  
; MOLECULE TYPE: peptide  
US-08-804-227C-10

Query Match 5.9%; Score 114; DB 2; Length 3724;

Best Local Similarity 23.7%; Pred. No. 0.17;

Matches 85; Conservative 48; Mismatches 139; Indels 86; Gaps 18;

QY 12 IEQKAVVKTDSVPELKEGTALVKVEAVAGNPTDKWKHIAIKIGPEGSILGCDIAGTVVK 71  
Db 2955 LDGLALVPAPDAEAP-LEPGQVRVAVRAAGVNFED-ALIALGMYPGEAEMGTGAGTVVE 3012  
QY 72 LGPNASTDLKVGDTGFGFVHGASQTDPKNGAFAYARVYPPFYKSNLTHSTADEISEGP 131  
Db 3013 VGPGV-TGVAVGDRVLGLWDGG-----LGPL-----CVADHRLAP 3047  
QY 132 VK---NFESAASLPVSLTTAGVSLCHLGSKMWHPTSTPQHTHPLLIWGATAVGQOLIQ 188  
Db 3048 VPDGWSYAQAASVPAPVLSAYYGLVTLAQLR-----PG-----BRVLHAAAGGVGMAAVQ 3098  
QY 189 VAKHINAYTKIVTVASKKHEKLLKSYGADDVFDYHDAGVIEQIKSKYPNLQHVDAVGSE 248  
Db 3099 IARHLGA--EVLATASPGKWDALRAMGITO-----DHLASS 3132  
QY 249 DSIPKAYKVT-ADSLPATLLEVPMTIESIPBEIRKDNVKIDITL-LYRASGOEILLGAT 306  
Db 3133 RTLDFAFTAGDG--TSRADVV---LNSLTKEF-----VDASGLLRPGGRFLELGT 3181  
QY 307 RFPASPEYHEATVKVFKFINPHLNN-----GDIHNMNKFVNSG-LDDVPALTEGK 357  
Db 3182 DV-RDPERIAAEHPGVRYRAFNLNEAGPDALGRLLRELMDLFAAGVHLPLPVVTHDVR 3238

US-08-804-227C-10

## RESULT 27

US-08-804-198-4  
; Sequence 4, Application US/08804198  
; Patent No. 5945320  
; GENERAL INFORMATION:  
; APPLICANT: Burgett, Stanley G.  
; APPLICANT: Kuhstoss, Stuart A.  
; APPLICANT: Rao, Nagaraja R.  
; APPLICANT: Richardson, Mark A.  
; APPLICANT: Rostock, Paul R., Jr.  
; TITLE OF INVENTION: PLATENOLIDE SYNTHASE GENE  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: PAUL R. CANTRELL 1138  
; STREET: LILLY CORPORATE CENTER  
; CITY: INDIANAPOLIS  
; STATE: IN  
; COUNTRY: USA  
; ZIP: 46285  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: Macintosh  
; OPERATING SYSTEM: Macintosh 7.0  
; SOFTWARE: Microsoft Word 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/804,198  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: CANTRELL, PAUL R.  
; REGISTRATION NUMBER: 36,470  
; REFERENCE/DOCKET NUMBER: P9113  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 317-276-3885  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3724 amino acids  
; TYPE: amino acid  
; TOPOLOGY: unknown  
; MOLECULE TYPE: peptide  
US-08-804-198-4

Query Match 5.9%; Score 114; DB 2; Length 3724;

Best Local Similarity 23.7%; Pred. No. 0.17;

Matches 85; Conservative 48; Mismatches 139; Indels 86; Gaps 18;

QY 12 IEQKAVVKTDSVPELKEGTALVKVEAVAGNPTDKWKHIAIKIGPEGSILGCDIAGTVVK 71  
Db 2955 LDGLALVPAPDAEAP-LEPGQVRVAVRAAGVNFED-ALIALGMYPGEAEMGTGAGTVVE 3012  
QY 72 LGPNASTDLKVGDTGFGFVHGASQTDPKNGAFAYARVYPPFYKSNLTHSTADEISEGP 131  
Db 3013 VGPGV-TGVAVGDRVLGLWDGG-----LGPL-----CVADHRLAP 3047  
QY 132 VK---NFESAASLPVSLTTAGVSLCHLGSKMWHPTSTPQHTHPLLIWGATAVGQOLIQ 188  
Db 3048 VPDGWSYAQAASVPAPVLSAYYGLVTLAQLR-----PG-----BRVLHAAAGGVGMAAVQ 3098  
QY 189 VAKHINAYTKIVTVASKKHEKLLKSYGADDVFDYHDAGVIEQIKSKYPNLQHVDAVGSE 248  
Db 3099 IARHLGA--EVLATASPGKWDALRAMGITO-----DHLASS 3132  
QY 249 DSIPKAYKVT-ADSLPATLLEVPMTIESIPBEIRKDNVKIDITL-LYRASGOEILLGAT 306  
Db 3133 RTLDFAFTAGDG--TSRADVV---LNSLTKEF-----VDASGLLRPGGRFLELGT 3181  
QY 307 RFPASPEYHEATVKVFKFINPHLNN-----GDIHNMNKFVNSG-LDDVPALTEGK 357  
Db 3182 DV-RDPERIAAEHPGVRYRAFNLNEAGPDALGRLLRELMDLFAAGVHLPLPVVTHDVR 3238

## RESULT 28

US-09-107-532A-5638



; Sequence 5638, Application US/09107532A  
; Patent No. 6583275  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; TITLE OF INVENTION: ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
; NUMBER OF SEQUENCES: 7310  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
; STREET: 100 Beaver Street  
; CITY: Waltham  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02354  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: CD-ROM ISO9660  
; COMPUTER: PC  
; OPERATING SYSTEM: <Unknown>  
; SOFTWARE: ASCII  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/107,532A  
; FILING DATE: 30-Jun-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/085,598  
; FILING DATE: 14 May 1998  
; APPLICATION NUMBER: 60/051571  
; FILING DATE: July 2, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ariniello, Pamela Deneke  
; REGISTRATION NUMBER: 40,489  
; REFERENCE/DOCKET NUMBER: GTC-012  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (781)893-5007  
; TELEFAX: (781)893-8277  
; INFORMATION FOR SEQ ID NO: 5638:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 200 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: YES  
; ORIGINAL SOURCE:  
; ORGANISM: Enterococcus faecium  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (8) LOCATION 1...200  
; SEQUENCE DESCRIPTION: SEQ ID NO: 5638:  
US-09-107-532A-5638  
  
Query Match 5.9%; Score 113.5; DB 4; Length 200;  
Best Local Similarity 26.3%; Pred. No. 0.0018;  
Matches 56; Conservative 37; Mismatches 75; Indels 45; Gaps 11;  
  
Qy 13 EGDKAVKVDVSVDELKEGTALVRVEAVAGNPTDVKHIAVYKIGPEGGIL-----GCD 64  
Db 18 EAQDVFEEIDAHPEVSGVGHVRVEIKAFSVNPDY---VALRLGEMKEIRTWKFPYVPGND 74  
  
Qy 65 IAGTVVKGPNASTDLKVGDTGFGFVHGASQTDPKNGAFAYARVYPLFYKSNLTHSTA 124  
Db 75 GAGLVTIGSDVIT-VHVGOR--VAHVAV-----GGTYGE--KVVLV-----S 112  
  
Qy 125 DEISEGPKV-NFESAASPLVSLTAGVSLCHHLGSKMEWHPSTPQHTHPLLLMGATVAG 183  
Db 113 AKVAKIPDKMSWEAEAGVTFGITA-YNLNLH-----TEIQPTDVTWILGASAVG 163  
  
Qy 184 QQLIQVAKHINATYKIVTVASKKHEKLLKSYGA 216  
Db 164 SSLIQLLHEKG--IRILTSASSKNEEKVTKFGA 194  
  
RESULT 29  
US-08-828-010-2  
; Sequence 2, Application US/08828010

; Patent No. 5955355  
; GENERAL INFORMATION:  
; APPLICANT: Chen, Hong  
; TITLE OF INVENTION: Eriemer, Nelson B.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR  
; TITLE OF INVENTION: THE DIAGNOSIS AND TREATMENT OF NEUROPSYCHIATRIC  
; TITLE OF INVENTION: DISORDERS  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10036/2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/828,010  
; FILING DATE: 27-MAR-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Coruzzi, Laura A  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 7853-068  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-790-9090  
; TELEFAX: 212-869-8864  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 300 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; FRAGMENT TYPE: internal  
; US-08-828-010-2

Query Match 5.8%; Score 111; DB 2; Length 300;  
Best Local Similarity 24.0%; Pred. No. 0.0061;  
Matches 80; Conservative 42; Mismatches 118; Indels 94; Gaps 17;  
  
Qy 52 YKIGPEGSILGCDIAGTVVKGPNASTDLKVGDTGFGFVHGASQTDPKNGAFAYARVYP 111  
Db 17 FDIGFEG-----IGEVVALGLSASARYTVG-----QAVAYMAP--GSPAEYTVVPA 60  
  
Qy 112 PLFYKSNLTHSTADEISEGPKVNF-SAASLPVSLTTAGVSLCHHLGSKMEWHPSTPQHT 170  
Db 61 SI-----ATPVPSEVKPEYLTLLVSGTTAVISL-KELGGLSE-----G 96  
  
Qy 171 HPLLWGGATACVQQLIQVAKHINATYKIVTVASKKHEKLLKSYGADDDVPDYHDAGVIEQ 230  
Db 97 KKVLTAAAGGTGQPAQMLSKAKACHV-IGTCSSEKSAFLKSLGCDRPNYKTEPVGTV 155  
  
Qy 231 IKSYP-NLQHVTDVGSSEDSPEAYKVTADSLPA-----AMFDLAVDALATKRLIVIGFISGYQTPTGLSPVKAG 210  
Db 156 LKQYEGVDVVYVESVGG-----YPA-----SQO-----EILLGATRFPSPEYH- 315  
  
Qy 276 SIPEEIRKONVKIDITLL-----YPA-----SQO-----EILLGATRFPSPEYH- 315  
Db 211 TLPKLLKKSASVQGFPLNHYLSKYQAAMSHLLMEVCVSGDLVCEVDLG-----DLSPEGRF 266  
  
Qy 316 ----EATVKVVKFINPHLNGDI-----HNMNKV 341  
Db 267 TGLSIFRAVNYMGMKNTGKIIVPELPHSVNSKL 300



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 10, 2004, 10:07:24 ; Search time 71 Seconds  
(without alignments)  
1050.121 Million cell updates/sec

Title: US-10-081-644-2

Perfect score: 1920

Sequence: 1 MSVPTTQKAVIEGDKAVK.....ALTEGIKGNKNVYVARL 370

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747907 seqs, 201509753 residues

Total number of hits satisfying chosen parameters: 747907

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 90 summaries

Database :

Published Applications AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pdb.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pdb.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pdb.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pdb.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pdb.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pdb.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pdb.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pdb.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pdb.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pdb.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pdb.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09D\_PUBCOMB.pdb.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pdb.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pdb.\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pdb.\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10D\_PUBCOMB.pdb.\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pdb.\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pdb.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1920	100.0	370	14	US-10-081-644-2
2	1065.5	55.5	376	14	US-10-081-644-4
3	1021.5	53.2	377	14	US-10-081-644-6
4	1015.5	52.9	368	14	US-10-081-644-8
5	266	13.9	318	15	US-10-137-036-66
6	265	13.8	322	9	US-09-815-242-10703
7	221	11.5	331	15	US-10-156-761-8740
8	207	10.8	308	15	US-10-156-761-8739
9	206	10.7	338	9	US-09-815-242-10944
10	201.5	10.5	315	15	US-10-156-761-14981
11	197.5	10.3	329	12	US-10-369-493-16579
12	197	10.3	331	15	US-10-156-761-8218
13	193.5	10.1	334	15	US-10-156-761-9593
14	193	10.1	343	12	US-10-369-493-102
15	186.5	9.7	328	12	US-10-369-493-11505

Sequence 12902, A  
Sequence 14785, A  
Sequence 16435, A  
Sequence 679, App  
Sequence 6825, Ap  
Sequence 8836, Ap  
Sequence 3394, Ap  
Sequence 5302, Ap  
Sequence 65, Appl  
Sequence 1893, Ap  
Sequence 14009, A  
Sequence 9450, Ap  
Sequence 3730, Ap  
Sequence 11902, A  
Sequence 69, Appl  
Sequence 53, Appl  
Sequence 4692, Ap  
Sequence 1245, Ap  
Sequence 20316, A  
Sequence 20590, A  
Sequence 21665, A  
Sequence 23606, A  
Sequence 11926, A  
Sequence 26, Appl  
Sequence 6782, Ap  
Sequence 677, App  
Sequence 72, Appl  
Sequence 12293, A  
Sequence 1878, Ap  
Sequence 5629, Ap  
Sequence 12251, A  
Sequence 2, Appl  
Sequence 10, Appl  
Sequence 1928, Ap  
Sequence 23095, A  
Sequence 4703, Ap  
Sequence 7461, Ap  
Sequence 8236, Ap  
Sequence 2023, Ap  
Sequence 20721, A  
Sequence 12263, A  
Sequence 5117, Ap  
Sequence 14, Appl  
Sequence 3785, Ap  
Sequence 20, Appl  
Sequence 20, Appl  
Sequence 355, App  
Sequence 10162, A  
Sequence 833, App  
Sequence 6285, Ap  
Sequence 10436, A  
Sequence 11395, A  
Sequence 11244, A  
Sequence 16582, A  
Sequence 52, Appl  
Sequence 53, Appl  
Sequence 1472, Ap  
Sequence 11164, A  
Sequence 7741, Ap  
Sequence 43, Appl  
Sequence 45, Appl  
Sequence 43, Appl  
Sequence 45, Appl  
Sequence 16399, A  
Sequence 14953, A  
Sequence 23041, A  
Sequence 20102, A  
Sequence 21322, A  
Sequence 21207, A  
Sequence 262, App  
Sequence 23134, A

89 114 5.9 350 12 US-10-369-493-22735 Sequence 22735, A  
90 114 5.9 1298 12 US-10-238-075-959 Sequence 959, App

## ALIGNMENTS

## RESULT 1

US-10-081-644-2  
; Sequence 2, Application US/10081644  
; Publication No. US20020192782A1  
; GENERAL INFORMATION:  
; APPLICANT: Yamamoto, Hiroaki  
; APPLICANT: Kimoto, No. US20020192782A1hiro  
; TITLE OF INVENTION: NOVEL ENONE REDUCTASES, METHODS FOR  
; TITLE OF INVENTION: PRODUCING SAME, AND METHODS FOR SELECTIVELY REDUCING A  
; TITLE OF INVENTION: CARBON-CARBON DOUBLE BOND OF AN ALPHA,BETA-UNSATURATED KETONE  
; TITLE OF INVENTION: USING THE REDUCTASES  
; FILE REFERENCE: 06501-100001  
; CURRENT APPLICATION NUMBER: US/10/081,644  
; PRIOR FILING DATE: 2002-02-21  
; PRIOR APPLICATION NUMBER: JP 2001-49363  
; PRIOR FILING DATE: 2001-02-23  
; NUMBER OF SEQ ID NOS: 29  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 370  
; TYPE: PRT  
; ORGANISM: Kluyveromyces lactis  
US-10-081-644-2

Query Match 100.0%; Score 1920; DB 14; Length 370;  
Best Local Similarity 100.0%; Pred. No. 3.7e-172;  
Matches 370; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSVPTTQKAVIIEGDKAVVKTDSVPELKEGTALVKVEAVAGNPTDMKHIAYKIGPEGSI 60  
DB 1 MSVPTTQKAVIIEGDKAVVKTDSVPELKEGTALVKVEAVAGNPTDMKHIAYKIGPEGSI 60  
QY 61 LGCDIAGTVVVKLGPNASTDLKVGDTGFGFVHGASQTDPKNGAFEAERYVPPFLYKSNLT 120  
DB 61 LGCDIAGTVVVKLGPNASTDLKVGDTGFGFVHGASQTDPKNGAFEAERYVPPFLYKSNLT 120  
QY 121 HSTADEISEGPKVKNFESAASLPVSLTTAGVSLCHLGSKMWEHPSTPOHTHPLINGCAT 180  
DB 121 HSTADEISEGPKVKNFESAASLPVSLTTAGVSLCHLGSKMWEHPSTPOHTHPLINGCAT 180  
QY 181 AVGGQLIQVAKHINAYTKIYTVVASKKHEKLLKSYGADDVFDYHDAGVIEQIKSKYPNLQH 240  
DB 181 AVGGQLIQVAKHINAYTKIYTVVASKKHEKLLKSYGADDVFDYHDAGVIEQIKSKYPNLQH 240  
QY 241 VIDAVGSEDSIPEAYKVTADSLPATLLEVPMTIESIPEIRKDNVKIDITLLYRASGQE 300  
DB 241 VIDAVGSEDSIPEAYKVTADSLPATLLEVPMTIESIPEIRKDNVKIDITLLYRASGQE 300  
QY 301 ILGATFPASPPEYHEATVVFVFINPHLNGDIHNMNIVKFSNGLDDVPALTEGIEK 360  
DB 301 ILGATFPASPPEYHEATVVFVFINPHLNGDIHNMNIVKFSNGLDDVPALTEGIEK 360  
QY 361 KNNKVVYVARL 370  
DB 361 KNNKVVYVARL 370

## RESULT 2

US-10-081-644-4  
; Sequence 4, Application US/10081644  
; Publication No. US20020192782A1  
; GENERAL INFORMATION:  
; APPLICANT: Yamamoto, Hiroaki  
; APPLICANT: Kimoto, No. US20020192782A1hiro  
; TITLE OF INVENTION: NOVEL ENONE REDUCTASES, METHODS FOR  
; TITLE OF INVENTION: PRODUCING SAME, AND METHODS FOR SELECTIVELY REDUCING A

; TITLE OF INVENTION: CARBON-CARBON DOUBLE BOND OF AN ALPHA,BETA-UNSATURATED KETONE  
; FILE REFERENCE: 06501-100001  
; CURRENT APPLICATION NUMBER: US/10/081,644  
; CURRENT FILING DATE: 2002-02-21  
; PRIOR APPLICATION NUMBER: JP 2001-49363  
; PRIOR FILING DATE: 2001-02-23  
; NUMBER OF SEQ ID NOS: 29  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 4  
; LENGTH: 376  
; TYPE: PRT  
; ORGANISM: Saccharomyces cerevisiae  
US-10-081-644-4

Query Match 55.5%; Score 1065.5; DB 14; Length 376;  
Best Local Similarity 54.0%; Pred. No. 1.2e-91;  
Matches 201; Conservative 59; Mismatches 109; Indels 3; Gaps 2;  
QY 2 SVPTTQKAVIIEGDKAVVKTDSVPELKEGTALVKVEAVAGNPTDMKHIAYKIGPEGSI 61  
DB 4 SIPETWKAIVIENGKAVVQDIPPELEEGFVLKTVAVAGNPTDMKHIYDFKIGPQCAL 63  
QY 62 GCDIAGTVVVKLGPNASTDLKVGDTGFGFVHGASQTDPKNGAFEAERYVPPFLYK--SN 118  
DB 64 GCDAGQIVKLGPNVDAARFAIGDYIYGVHIGASVRFPSNGAFEAESAISSETAYKPARE 123  
QY 119 LTHSTADEISEGPKVKNFESAASLPVSLTTAGVSLCHLGSKMWEHPSTPOHTHPLINGG 178  
DB 124 FRLCGKDKLPEGPKVKSLEGAVALPSVSLTTAGMLTHTSFLGDMTWKPSKAQRDPILFWGG 183  
QY 179 ATAVGQOLIQVAKHINAYTKIYTVVASKKHEKLLKSYGADDVFDYHDAGVIEQIKSKYPNL 238  
DB 184 ATAVGQMLQLAKLNGFSGKIYVASKKHEKLLKSYGADDELFDYHDADVIEQIKKYNKI 243  
QY 239 QHVIDAVGSEDSIPEAYKVTADSLPATLLEVPMTIESIPEIRKDNVKIDITLLYRASG 298  
DB 244 PYLVDCVSNTEITQQVYKCAADDLDATVVQLTVLTKEDIKEEDRRQNVSIETFLYLG 303  
QY 299 QEILGATFPASPPEYHEATVVFVFINPHLNGDIHNMNIVKFSNGLDDVPALTEGIEK 358  
DB 304 NDVPFGTFLPADPEYKEAAIKFIKPKINDGETHHIPVKVYKNGLDDIPQLLDDIKH 363  
QY 359 GKNKVVYVARL 370  
DB 364 GRNKGKLVAVL 375

## RESULT 3

US-10-081-644-6  
; Sequence 6, Application US/10081644  
; Publication No. US20020192782A1  
; GENERAL INFORMATION:  
; APPLICANT: Yamamoto, Hiroaki  
; APPLICANT: Kimoto, No. US20020192782A1hiro  
; TITLE OF INVENTION: NOVEL ENONE REDUCTASES, METHODS FOR  
; TITLE OF INVENTION: PRODUCING SAME, AND METHODS FOR SELECTIVELY REDUCING A  
; TITLE OF INVENTION: CARBON-CARBON DOUBLE BOND OF AN ALPHA,BETA-UNSATURATED KETONE  
; TITLE OF INVENTION: USING THE REDUCTASES  
; FILE REFERENCE: 06501-100001  
; CURRENT APPLICATION NUMBER: US/10/081,644  
; CURRENT FILING DATE: 2002-02-21  
; PRIOR APPLICATION NUMBER: JP 2001-49363  
; PRIOR FILING DATE: 2001-02-23  
; NUMBER OF SEQ ID NOS: 29  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 6  
; LENGTH: 377  
; TYPE: PRT  
; ORGANISM: Saccharomyces cerevisiae  
US-10-081-644-6

Query Match 53.2%; Score 1021.5; DB 14; Length 377;

Best Local Similarity 51.7%; Pred. No. 1.6e-87; Mismatches 193; Conservative 67; Mismatches 110; Indels 3; Gaps 2;

QY 1 MSVPTTQKAVIIEGDKAVVKTDSVPELKEGTALVKVEAVAGNPTDWHKHYKIGEGSI 60  
DB 4 VAIPETMKAVIIEGDKAVVKEGIPPELEBEGFVLKTLAVAGNPTDWHKHYKIGEGSI 63  
QY 61 LGCDDIAGTVVKGPNAS--TDLKVGDTGFGFVHGASQTDPKNGAFAYARVYPLFYKS-- 117  
DB 64 LGCDAAGQIVKLGPAVNPXDFSGDIYGFHIGSSVRFPFNGAFAYSAISTVWAYKSPN 123  
QY 118 NLTHTADISBGPVKNFSAASLPVSLTTAGVSLCHLGSKMEWHPSTPQTHPLLIWG 177  
DB 124 ELKFLGEDVLPAGVRSLEGVATIPVSLTTAGVLVLTYNLGLDKMEPSTPQKGPILLWG 183  
QY 178 GATVAGQQLIQVAKHINAYTKVTVASKKHEKLLKSYGADDVFDYHDAGVIEQIKSKYPN 237  
DB 184 GATVAGQQLIQVAKHINAYTKVTVASKKHEKLLKSYGADDVFDYHDAGVIEQIKSKYPN 243  
QY 238 LQHVDAVGSBDSIPAYKVTADSLPATLLEVPMTIESIPEIRKDNVVKIDITLLYRAS 297  
DB 244 ISYLVDCVANQDTLQOYKCAADKQDATIVELKNLTEENVKKNRQNTVIDIRLYSIG 303  
QY 298 GQIILLGATRFASPEYHEATVKFVFNPHLNGDIHNMNKKVFSNGLDDVPALTEGIG 357  
DB 304 GHEVPFGNITLPADSEARKAAIKFIPKINDGQIRHPVRVYKNGLCVPHILKDIK 363  
QY 358 EGKKNVKKVAVL 370  
DB 364 YKNSGEKLVAVL 376

RESULT 4  
US-10-081-644-8  
; Sequence 8, Application US/10081644  
; Publication No. US20020192782A1  
; GENERAL INFORMATION:  
; APPLICANT: Yamamoto, Hiroaki  
; TITLE OF INVENTION: NOVEL ENONE REDUCTASES, METHODS FOR  
; TITLE OF INVENTION: PRODUCING SAME, AND METHODS FOR SELECTIVELY REDUCING A  
; TITLE OF INVENTION: CARBON-CARBON DOUBLE BOND OF AN ALPHA, BETA-UNSATURATED KETONE  
; TITLE OF INVENTION: USING THE REDUCTASES  
; FILE REFERENCE: 06501-100001  
; CURRENT APPLICATION NUMBER: US/10/081,644  
; PRIOR FILING DATE: 2002-02-21  
; PRIOR APPLICATION NUMBER: JP 2001-49363  
; NUMBER OF SEQ ID NOS: 29  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 8  
; LENGTH: 368  
; TYPE: PRT  
; ORGANISM: Saccharomyces cerevisiae  
US-10-081-644-8

Query Match 52.9%; Score 1015.5; DB 14; Length 368;  
Best Local Similarity 53.0%; Pred. No. 5.8e-87;  
Matches 194; Conservative 60; Mismatches 109; Indels 3; Gaps 2;

QY 8 KAVIIEGDKAVVKTDSVPELKEGTALVKVEAVAGNPTDWHKHYKIGEGSILGCDIAG 67  
DB 2 KAVIIEGDKAVVKEGPIPELEBEGFVLKTLAVAGNPTDWHKHYKIGEGSILGCDIAG 61  
QY 68 TVVKGPNAS--TDLKVGDTGFGFVHGASQTDPKNGAFAYARVYPLFYKS--NLTHSTA 124  
DB 62 QIVKLGPAVDPKDFSGDIYGFHIGSSVRFPFNGAFAYSAISTVWAYKSPNELFLGE 121  
QY 125 DETSEBGPVKNFSAASLPVSLTTAGVSLCHLGSKMEWHPSTPQTHPLLIWGATVAGQ 184  
DB 122 DVLVLPAGVRSLEGVATIPVSLTTAGVLVLTYNLGLDKMEPSTPQKGPILLWGATVAGQ 181  
QY 185 QLTQVAKHINAYTKVTVASKKHEKLLKSYGADDVFDYHDAGVIEQIKSKYPNLQHVDA 244

DB 182 SLIQLANKLNGFTKIIVVASRKHKKLKEYGADQLFDYHDIDVVEQIKHKNYNNISYLVDC 241  
QY 245 VGSBDSIPBEAYKVTADSLPATLLEVPMTIESIPEIRKDNVVKIDITLLYRASGOBILG 304  
DB 242 VANQTLQOYKCAADKQDATIVELKNLTEENVKKNRQNTVIDIRLYSIGGHEVPPG 301  
QY 305 ATRFPASPEYHEATVKFVFNPHLNGDIHNMNKKVFSNGLDDVPALTEGIGKGNKVV 364  
DB 302 GITFPADPEARAAATEFVKFNPKNISDQIHHIPARVYKNGLYDVPRILEDIKIGNSGE 361  
QY 365 KYVAVL 370  
DB 362 KLVAVL 367

RESULT 5  
US-10-137-036-66  
; Sequence 66, Application US/10137036  
; Publication No. US20030101478A1  
; GENERAL INFORMATION:  
; APPLICANT: Perera, Ranjan  
; APPLICANT: Rice, Stephen  
; APPLICANT: Bagleton, Clare  
; APPLICANT: Lasham, Annette  
; APPLICANT: Wood, Marion  
; APPLICANT: Visser, Elizabeth  
; TITLE OF INVENTION: Compositions and Methods for the  
; TITLE OF INVENTION: Modification of Gene Expression  
; FILE REFERENCE: 11000.1036c4  
; CURRENT APPLICATION NUMBER: US/10/137,036  
; PRIOR FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: PCT/NZ 01/00115  
; PRIOR FILING DATE: 2001-06-20  
; PRIOR APPLICATION NUMBER: U.S. No. US20030101478A1 09/724,624  
; PRIOR FILING DATE: 2000-11-28  
; PRIOR APPLICATION NUMBER: U.S. No. US20030101478A1 09/598,401  
; PRIOR FILING DATE: 2000-06-20  
; PRIOR APPLICATION NUMBER: PCT/NZ00/00018  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: U.S. No. US20030101478A1 60/146,591  
; PRIOR FILING DATE: 1999-07-30  
; PRIOR APPLICATION NUMBER: U.S. No. US20030101478A1 09/276,599  
; PRIOR FILING DATE: 1999-03-25  
; NUMBER OF SEQ ID NOS: 143  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 66  
; LENGTH: 318  
; TYPE: PRT  
; ORGANISM: Eucalyptus grandis  
US-10-137-036-66

Query Match 13.9%; Score 266; DB 15; Length 318;  
Best Local Similarity 33.6%; Pred. No. 1.8e-16;  
Matches 91; Conservative 31; Mismatches 93; Indels 56; Gaps 11;

QY 3 VPTTOKAVIIEGDKAVV---KTDVSVPELKEGTALVKVEAVAGNPTDWHKHYKIGPE 57  
DB 7 IPTKMKAVYRHRGDNVANVLGDDPELKEGQVVLVKVLAALNPIDTARVKGIVKLP 66  
QY 58 G----SILGCDIAGTVVKGPNAS--TDLKVGDTGFGFVHGASQTDPKNGAFAYARVYPL 113  
DB 67 GFSLPVAVPGYDLGAVVVKVGVK--ELKVGDEYGVFMPHAK---KDGTLAEYAAVEESP 121  
QY 114 FYKSNLTHSTADEISEGPVK--NPFESAASLPVSLTT-----AGVSLCHLGSKMEWHP 165  
DB 122 -----LALPKYKLFGEAASLPVVIQTAYGGLERAGLS----- 154  
QY 166 TPQHTHPLLIWGATVAGQQLIQVAKHINAYTKVTVASKKHEKLLKSYGADDVFDYHDA 225  
DB 155 ---HGKSLVLVGGAGVGTLLIQLAKEVFGASRVAATSTGKLELLKSLGADLAIDYTKV 211  
QY 226 GVIEQIKSKYPNLQHVDAVGSBDSIPEAYK 256

```

RESULT 7
US-10-156-761-8740
; Sequence 8740, Application US/10156761.
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156.761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 8740
; LENGTH: 331
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-8740

Query Match      11.5%; Score 221; DB 15; Length 331;
Best Local Similarity 24.7%; Pred. No. 3.3e-12;
Matches 94; Conservative 60; Mismatches 131; Indels 96; Gaps 16

Qy 10 VIIEGKAVVKTIVSVPELKEGTALVKVEAVAGNPTDWK-----HIAYKIGEGSIL 61
Db 5 VVTKYEPLQEADVPEVTGEHDVLVRVEAAGLNPLDEKIRAGEFKQLIPYKL---PLIL 61

Qy 62 GCDIAGTVVKLGNASTDLKAVGTGTFGVGASQTDPKNGAFABYARVYPLFYKSNLTH 121
Db 62 GNDVAGTVISVG-TAVRGFKPGDEVYARPH-----QDRIGTFABRIAI----- 103

Qy 122 STADEISEGPKV-NFRSAALPVSLLTAGVSLCH---HLGSKMEWHPSTPOHTPLLIW 176
Db 104 -AEDDLALKPASISMEENGSLPLAALTAWQALVERGKVRPGQK-----VLIIH 149

Qy 177 GGATVAGQQLIQVAKHINATKIVTVASKKHEKLLKSYGADDPYDHDAGVIEQISKYP 236
Db 150 AGAGGVGSAIQAALHGA--SVATTASGNSADFVALGADTVIDYRSQD-FEQLLTGY- 205

Qy 237 NLQHVTDVAGSEDSISPEAYKV-----TADSLPATLLEVVVPMTIESIPEEI 281
Db 206 --DLVLDLSLGE-NLEKSLRVLPKGGKAIAGIAGPPDPAFAREAGLNPLRLAVAGMSGKI 262

Qy 282 RKNQVKIDIT---LLVRASQOEILLGATRPFPASPEYHEATVKFVKFINPHLNNGDIHHMN 338
Db 263 RRQAKKLGVTYEFLEWRASQDL-----RQITLLDQGVVRPVV 301

Qy 339 IKVFSNGLDVPALTEGIEKEG 359
Db 302 GKVV--GFQDOTQALQSLSQG 320

```

RESULT 8  
US-10-156-761-8739  
; Sequence 8739, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

```
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIORITY APPLICATION NUMBER: JP 2001-204089
PRIORITY FILING DATE: 2001-05-30
PRIORITY APPLICATION NUMBER: JP 2001-272697
PRIORITY FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 8739
LENGTH: 308
TYPE: PRT
ORGANISM: Streptomyces avermitilis
US-10-156-761-8739

Query Match 10.8%; Score 207; DB 15; Length 308;
Best Local Similarity 26.9%; Pred. No. 6.2e-11;
Matches 98; Conservative 43; Mismatches 111; Indels 112; Gaps 19;

QY 8 KAVIIE---GDKAVKTDVSVPELKEGTALVKVEAVAGNPTDWHIAIKTGPESIL--- 61
DB 2 RAVVVEQWGGPEVLVEREVARPEPLNEVLVRVHAAGVNPDPKTRA-----SCALIEMG 56

QY 62 -----GCDIAGTVVKLGPNASTDLKVGDTGFGFVHGASQTDPKNGAFAY-----ARVYPP 112
DB 57 EVPAVGMDSVGTVEAVGPGVGM-FRPGDE-----VYGMPLFPQAGAYAYVAPARHLAP 111

QY 113 LFYKSNLTHSTADEISEGPKNFESAAASLPVSLTTAGVSLCHLHGSKMWHHPSTPQHTHP 172
DB 112 --KPNLTH-----VQAAALPLAALTAWQALVDTAGVRA-----GER 146

QY 173 LLIWGGATVGOQLIQVAKHINAYTKIVTVASKKHEKLLKSYGADVDYHDAGVIEQIK 232
DB 147 VLVAAGAGVGHVLAQIAKARGAYV-IGTASAGKHD-LLRQLGADEVIDYRTV----- 197

QY 233 SKYPNLOHVDAVGSSEDSIPEAYKVTADSLPATLLEVVVPTIESIPEIRKDNVKDITL 292
DB 198 -----RFDVAGDVDP-----WLDGLGG-----QNAERSLTV 224

QY 293 LYRASGOEILL-GATRPASPVEHEATVKFKFINP-HLANNGDIIHNNIKVFSNGLDVDP 350
DB 225 L-RPGGLTLTLPDGDVDPADVPGHVRAVWML--VEPDHL-----GLREIA 266

QY 351 ALTE 354
DB 267 ALAE 270

RESULT 9
US-09-815-242-10944
Sequence 10944, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Hasebeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
TITLE OF INVENTION: Prokaryotes
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIORITY APPLICATION NUMBER: 60/191,078
PRIORITY FILING DATE: 2000-03-21
PRIORITY APPLICATION NUMBER: 60/206,848
PRIORITY FILING DATE: 2000-05-23
PRIORITY APPLICATION NUMBER: 60/207,727
PRIORITY FILING DATE: 2000-05-26
PRIORITY APPLICATION NUMBER: 60/242,578
PRIORITY FILING DATE: 2000-10-23

FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIORITY APPLICATION NUMBER: JP 2001-204089
PRIORITY FILING DATE: 2001-05-30
PRIORITY APPLICATION NUMBER: JP 2001-272697
PRIORITY FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 8739
LENGTH: 308
TYPE: PRT
ORGANISM: Streptomyces avermitilis
US-10-156-761-8739

Query Match 10.8%; Score 207; DB 15; Length 308;
Best Local Similarity 26.9%; Pred. No. 6.2e-11;
Matches 98; Conservative 43; Mismatches 111; Indels 112; Gaps 19;

QY 8 KAVIIE---GDKAVKTDVSVPELKEGTALVKVEAVAGNPTDWHIAIKTGPESIL--- 61
DB 2 RAVVVEQWGGPEVLVEREVARPEPLNEVLVRVHAAGVNPDPKTRA-----SCALIEMG 56

QY 62 -----GCDIAGTVVKLGPNASTDLKVGDTGFGFVHGASQTDPKNGAFAY-----ARVYPP 112
DB 57 EVPAVGMDSVGTVEAVGPGVGM-FRPGDE-----VYGMPLFPQAGAYAYVAPARHLAP 111

QY 113 LFYKSNLTHSTADEISEGPKNFESAAASLPVSLTTAGVSLCHLHGSKMWHHPSTPQHTHP 172
DB 112 --KPNLTH-----VQAAALPLAALTAWQALVDTAGVRA-----GER 146

QY 173 LLIWGGATVGOQLIQVAKHINAYTKIVTVASKKHEKLLKSYGADVDYHDAGVIEQIK 232
DB 147 VLVAAGAGVGHVLAQIAKARGAYV-IGTASAGKHD-LLRQLGADEVIDYRTV----- 197

QY 233 SKYPNLOHVDAVGSSEDSIPEAYKVTADSLPATLLEVVVPTIESIPEIRKDNVKDITL 292
DB 198 -----RFDVAGDVDP-----WLDGLGG-----QNAERSLTV 224

QY 293 LYRASGOEILL-GATRPASPVEHEATVKFKFINP-HLANNGDIIHNNIKVFSNGLDVDP 350
DB 225 L-RPGGLTLTLPDGDVDPADVPGHVRAVWML--VEPDHL-----GLREIA 266

QY 351 ALTE 354
DB 267 ALAE 270

RESULT 9
US-09-815-242-10944
Sequence 10944, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Hasebeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
TITLE OF INVENTION: Prokaryotes
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIORITY APPLICATION NUMBER: 60/191,078
PRIORITY FILING DATE: 2000-03-21
PRIORITY APPLICATION NUMBER: 60/206,848
PRIORITY FILING DATE: 2000-05-23
PRIORITY APPLICATION NUMBER: 60/207,727
PRIORITY FILING DATE: 2000-05-26
PRIORITY APPLICATION NUMBER: 60/242,578
PRIORITY FILING DATE: 2000-10-23

FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIORITY APPLICATION NUMBER: JP 2001-204089
PRIORITY FILING DATE: 2001-05-30
PRIORITY APPLICATION NUMBER: JP 2001-272697
PRIORITY FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 10944
LENGTH: 338
TYPE: PRT
ORGANISM: Enterococcus faecalis
US-09-815-242-10944

Query Match 10.7%; Score 206; DB 9; Length 338;
Best Local Similarity 28.1%; Pred. No. 8.9e-11;
Matches 83; Conservative 39; Mismatches 87; Indels 86; Gaps 16;

QY 7 QKAVIIEGKAVVKTVDVSVPELKEGTALVKVEAVAGNPTDWHIAIKTGPESIL--- 58
DB 12 QKELATE-----EVPLPTIHDNDVLVKIIAASINPIDLTKDQKVKMLNMQM---P 60

QY 59 SILGCDIAGTVVKLGPNASTDLKVGDTGFGFVHGASQTDPKN---GAPAEYARV---YPPLF 114
DB 61 LILGSDFAGIVSVGKKVQ-NPRLGDAVYGRV-----PKNRVGTFABYIADVDAQAAVAM 112

QY 115 YKSNLTHSTADEISEGPKNFESAAASLPVSLTTAGVSLCHLHGSKMWHHPSTPQHTHP 174
DB 113 KPNLTH-----FEENAAIPL-----VGLTSQALHDMNVQPG-----QKVL 149

QY 175 IMGATVAVGOQLIQVAKHINAYTKIVTVASKKHEKLLKSYGADVDYHDAGVIEQIKSK 234
DB 150 IQAGSGGIGTIAIQLAKLAGAY--VATTTSSKNKEWQALGADVIDYRTON-FEEVLSD 206

QY 235 YPNLOHVDAVGSSEDSIPEAYKVTADSLPATLLE-----VVP-----MTIESIEE 280
DB 207 Y---DYVFDTMG-----GTILEKAFSVVYKPGQKVVTLTSGIPNE 241

RESULT 10
US-10-156-761-14981
Sequence 14981, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIORITY APPLICATION NUMBER: JP 2001-204089
PRIORITY FILING DATE: 2001-05-30
PRIORITY APPLICATION NUMBER: JP 2001-272697
PRIORITY FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 14981
LENGTH: 315
TYPE: PRT
ORGANISM: Streptomyces avermitilis
US-10-156-761-14981

Query Match 10.5%; Score 201.5; DB 15; Length 315;
Best Local Similarity 25.9%; Pred. No. 2.1e-10;
Matches 80; Conservative 49; Mismatches 125; Indels 55; Gaps 15;

QY 1 MSVPTTQKAV---IIEGDKAVKTDVSVPELKEGTALVKVEAVAGNPTDWHIAIKT---IG 55
DB 1 MSTVNTMRAISQDVLGGPEVLKEVTERPAPRNEVLVRVRAAGVNPVDWKHRTATGFLG 60
```







Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; PRIOR FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 12902  
; LENGTH: 353  
; TYPE: PRT  
; ORGANISM: Aspergillus nidulans  
US-10-369-493-12902

Query Match 9.5%; Score 182; DB 12; Length 353;  
Best Local Similarity 23.6%; Pred. No. 1.7e-08;  
Matches 90; Conservative 57; Mismatches 149; Indels 86; Gaps 15;

QY 3 VPTOKAVIIEGDKAVVKTDSV-----PELKE--GTALVKVEAVAGNPTDKW---HIA 51  
DB 9 IPPTMKSLVYSTQPSVISNLSFTASAPSPPLPSYANQLLIKVTLSNPADHKVPQHAT 68  
QY 52 YKIGPEGSL-----GCDIAGTVVVKLGPNASTDLKVGDTGFGFVHGASQTDPKNGAFA 104  
DB 69 IPTGGRTLICGLPASGLDFAGKIVSVHPKNRTEPEGQ-----LVVGLARPRTEGTTG 124  
QY 105 EYARVYPLPYKSNLTHSTDAETISEGPKVPFESAAISLPVSLTAGVSLCHHLG-SKWEWH 163  
DB 125 EYI-----LADANDVAH-----LPEG--VSVDAAACLGAVRTAYOSLKNYLDLSKFE-G 171  
QY 164 PSTPOHPTLLIIGGATAVGGQLIQVAKHINAYTKIVTVASKKHEKLLKSYGADVDVYH 223  
DB 172 GSGR-----VFINGSGGGCVFALQIAMLGC--QVTTCSGRNIELVRDLGADEIIDYT 225  
QY 224 DAGVIEQIKSNLQHVDAVGSSEDSIPEAYKVTADSLPATLLEVVPMPTIESIPEIRK 283  
DB 226 TTNVTELTKLGLVYDVIDHIG-----LPG----- 251  
QY 284 DNKIDITLYRAGSQBILGATRFPSPYHEATVKFVFINPHLNGDIHMHNIKVPFS 343  
DB 252 -NLVYAECHHFKPSGVWVQGA-----GSILTAFWRAITPRFLLGGRRWFVPLMMA 301  
QY 344 NGLDDVPALTEGIKEGKNVK 365  
DB 302 NSKEDLVEGNLUKEGKLRVK 323

RESULT 17  
US-10-369-493-14785  
; Sequence 14785, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 14785

Query Match 9.3%; Score 179; DB 12; Length 276;  
Best Local Similarity 26.2%; Pred. No. 2.3e-08;  
Matches 85; Conservative 49; Mismatches 123; Indels 68; Gaps 17;

QY 60 ILGCDIAGTVVVKLGPNASTDLKVGDTGFGFVHGASQTDPKNGAFAYARVYPLPYKSNL 119  
DB 5 ILGNDFSGVIVKVGSGV-TRFKVGDALY-----ARPKNKIGCTFAEYIAIH----- 49  
QY 120 THSTDAETISEGPKVN--FESAASLPVSLTAGVSL--CHILGSKWEWHPTQHTHPLLI 175  
DB 50 -----EDDIATLKP-KNLSFEESAISPLVGLTSYQALHDIMHL-----QKQKILI 93  
QY 176 WGGATAVGGQLIQVAKHINAYTKIVTVASKKHEKLLKSYGADVDVYHAGVIEQIKSKY 235  
DB 94 HAGSGVGTFAIQALAKINGA--TVTTTASEAGSLVKSGLGADQIINYK-----TEFEEIL 147  
QY 236 PNLQHVDAVGSSEDSIPEAYKVTADSLPATLLEVVPMPTIESIPEIRKDNKIDITLYR 295  
DB 148 KNYDAVFDITIG-TTLEKSFNIIKSG--GNIVSVSGMPNARPKBFGSGFFK---TLLFS 201

LENGTH: 333  
; TYPE: PRT  
; ORGANISM: Agrobacterium tumefaciens  
US-10-369-493-14785

Query Match 9.5%; Score 181.5; DB 12; Length 333;  
Best Local Similarity 28.3%; Pred. No. 1.8e-08;  
Matches 84; Conservative 37; Mismatches 117; Indels 59; Gaps 12;

QY 8 KAVIIEGDKAVK-----TDSVPKELKEGTALVKVEAVAGNPTDKW-----HIAYKI 54  
DB 2 KAFV--DKYKKKGLPLADMPNPGVIGANDVLRIHATAINLLDSKVRDGEFKFLPLPYR- 58  
QY 55 GPEGSTLGCIDIAGTVVVKLGPNASTDLKVGDTGFGFVHGASQTDPKNGAFAYARVYPLP 114  
DB 59 --PPFILGHDLAGTVIRVGANV-RQFKTGDEVF-----ARPRDHRVGTFAEMIAV----- 105  
QY 115 YKSNLTHSTDAETISEGPKVPNFESAASLPVSLTAGVSLCH-----HLGSKWEWHPTQHT 170  
DB 106 -----DAADLALKPTSLSMQAASIPLVGLTAQALIEVGVKSGQK----- 147  
QY 171 HPELLWGGATAVGGQLIQVAKHINAYTKIVTVASKKHEKLLKSYGADVDVYHAGVIEQ 230  
DB 148 --VFIQAGSGGVGTFAIQALAKHLGA--TVATTTSAANAEVKSGLGADVVIDYKTQD-FEQ 202  
QY 231 IESKYPNLQHVDAVGSSEDS---IPEAYKVTADSLPATLLEVVPMPTIESIPEIRKD 284  
DB 203 VLSGVDVLNLSQDAKTLEKSLNVLPRGGKLIISISGPPDVAFARSULKIESAPAFACRON 259

RESULT 18  
US-10-369-493-16435  
; Sequence 16435, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 16435  
; LENGTH: 276  
; TYPE: PRT  
; ORGANISM: Bacillus thuringiensis  
US-10-369-493-16435

Query Match 9.3%; Score 179; DB 12; Length 276;  
Best Local Similarity 26.2%; Pred. No. 2.3e-08;  
Matches 85; Conservative 49; Mismatches 123; Indels 68; Gaps 17;

QY 60 ILGCDIAGTVVVKLGPNASTDLKVGDTGFGFVHGASQTDPKNGAFAYARVYPLPYKSNL 119  
DB 5 ILGNDFSGVIVKVGSGV-TRFKVGDALY-----ARPKNKIGCTFAEYIAIH----- 49  
QY 120 THSTDAETISEGPKVN--FESAASLPVSLTAGVSL--CHILGSKWEWHPTQHTHPLLI 175  
DB 50 -----EDDIATLKP-KNLSFEESAISPLVGLTSYQALHDIMHL-----QKQKILI 93  
QY 176 WGGATAVGGQLIQVAKHINAYTKIVTVASKKHEKLLKSYGADVDVYHAGVIEQIKSKY 235  
DB 94 HAGSGVGTFAIQALAKINGA--TVTTTASEAGSLVKSGLGADQIINYK-----TEFEEIL 147  
QY 236 PNLQHVDAVGSSEDSIPEAYKVTADSLPATLLEVVPMPTIESIPEIRKDNKIDITLYR 295  
DB 148 KNYDAVFDITIG-TTLEKSFNIIKSG--GNIVSVSGMPNARPKBFGSGFFK---TLLFS 201







Db 100 ANTDMQAASLPLVALTA-----WQVLVETARLQKQKVLIIHAGSGGVCTVA 146  
Qy 187 IQVAKHINAYTKIIVTVASKHEKLLKSYGADDDVFDY-----HDAGVIEQIKSKYPN 237  
Db 147 IQLAKHILGAF--VATTSTANVEWVKALGADLVIDYTDQDFASVLHDYDV-----194  
Qy 238 LQHVIVDVGSDSIPAEYKV-----TADSLPAT-----LLEVPVMTIESIP 278  
Db 195 ---VLSLGA-DVLEKSLKVLKPGQLISISGPTTVQFAREHGLSWLPQVMRLLSLIGIR 250  
Qy 279 EEIRKONWKIDITLLYRASGOEI 301  
Db 251 RKARKQDVSFAFVFM-RANGAQL 272

## RESULT 27

US-10-156-761-9450  
; Sequence 9450, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 9450  
; LENGTH: 445  
; TYPE: PRT  
; ORGANISM: Streptomyces avermitilis  
US-10-156-761-9450

Query Match 7.8%; Score 150; DB 15; Length 445;  
Best Local Similarity 21.3%; Pred. No. 2.6e-05;  
Matches 93; Conservative 52; Mismatches 142; Indels 150; Gaps 17;  
Qy 1 MSVPTTKAVIIIGD-----KAVVKTDSVPELKEGTALVKVEAVAGNP 44  
Db 21 LPLPDSYRAITVHKDETEMFAGLSTRDKPRKSHLDDVPVPELGPCEALVAVMASSVNY 80  
Qy 45 TD-WKHAYKIGPEG-----SILGCDIAGTVVKLGPN-----76  
Db 81 NSVWTSIFEPVSTFNFLERYGRISDLKSRHDLPHYIIGSLAGVLRTPGVNSWKPGE 140  
Qy 77 -----STDLKVGDTGFGFVHGASQTPKN-----GAFAYEARVYPPPLFVKNLTH 121  
Db 141 VVAHCLSVLESSDG-----HNDTMLDPEORINGCFETNFGGLAEIALV-----KSNQLM 189  
Qy 122 STADEISEGPVKPNFESAASLPVSLTTAGVSLCHLHLSKMEWHPSTPOHTHPLIHWGATA 181  
Db 190 PKPDHLS-----WEEAAPGLVNSTAYRQLVSRNGAGMKQGDN-----VLIWGASGG 236  
Qy 182 VQOOLIQVAKHINAYTKIIVTVASKHEKLLKSYGADDDVFDYHDAGVIEQIKSKYPNLOHV 241  
Db 237 LGSYATQFALAGGA-NPICVVSSEQKADICRSMGAEAIIIDRN-----277  
Qy 242 IDAVGSDSIPAEYKVTADSLPATLLEVPVMTIESIPEEIRKONVKIDITLLYRASGOEI 301  
Db 278 -----AEGYKFWKDETTQD-----PKEWKRFGKRIREFTGGEIDIVFEHFGRE- 321  
Qy 302 LLGATRF-----PASPEYHEATVKFV-----KFINPHLNN-----331  
Db 322 TFGASVYVTRKGGTITTCASTSGYMEYDNRKYLWMSLKRIGSHFANYREAWEANRLVAK 381

Qy 332 GDIIHMNIKIVFNSGLDD 348  
Db 382 GKIHPTLSKVYS--LED 396  
RESULT 28  
US-09-738-626-3730  
; Sequence 3730, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIRO  
; APPLICANT: OCHIAI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO  
; APPLICANT: TATEISHI, NAKO  
; APPLICANT: SENOH, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OZAKI, AKIO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT APPLICATION NUMBER: US/09/738,626  
; CURRENT FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 99/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: JP 00/280988  
; PRIOR FILING DATE: 2000-08-03  
; NUMBER OF SEQ ID NOS: 7059  
; SOFTWARE: PatentIn ver. 3.0  
; SEQ ID NO 3730  
; LENGTH: 318  
; TYPE: PRT  
; ORGANISM: Corynebacterium glutamicum  
US-09-738-626-3730

Query Match 7.8%; Score 149; DB 10; Length 318;  
Best Local Similarity 27.0%; Pred. No. 1.9e-05;  
Matches 67; Conservative 33; Mismatches 96; Indels 52; Gaps 11;  
Qy 16 KAVVKT-----DVSVPKELKEGTALVKVEAVAGNPTDWKHIA--YKIGPEG--I 60  
Db 2 QAIQVTEEKVTASLELQEVPPVTLKPGVLEVKVKAAGVNRADLLQTOGNYPV-PAGASEI 60  
Qy 61 LGCDIAGTVVKLGPNASTDLKVGDTGFGFVHGASQTPDKNGAFAYEARVYPPPLFYKSNLT 120  
Db 61 LGLECAGVIVDAGDTGQT---VGQEVACLLTG-----GGVAVQVAV-----P 99  
Qy 121 HSTADSEISGPVKPNFESAASLPVSLTTAGVSLCHLHLSKMEWHPSTPOHTHPLIHWGAT 180  
Db 100 EGQLMPEPEG--YSFVEAASIVEVACTVMSNIGMLAQKE-----DTFLIHGAG 148  
Qy 181 AVQOOLIQVAKHINAYTKIIVTVASKHEKLLKSYGADDDVFDYHDAGVIEQIKSKYPNLOH 240  
Db 149 GIGTFAIQMGKALGV-TVAVTAGSTEKTKCNLGLADILINYKEEDFAEVLKNNK---ADV 204  
Qy 241 VIDAVGSE 248  
Db 205 ILDIIGAK 212  
RESULT 29  
US-10-156-761-11902  
; Sequence 11902, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI

```
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 11902
; LENGTH: 326
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-11902

Query Match      7.7%; Score 148.5; DB 15; Length 326;
Best Local Similarity 26.5%; Pred. No. 2.2e-05;
Matches 69; Conservative 37; Mismatches 99; Indels 55; Gaps 13;

Qy 14 GDKAVKTVDSVPELKEGTALVKEAVAGNPTD-WKHIAVKIGPEGS--ILGCDIAGTVV 70
Db 11 GPEALVWTEVPDEPGEGEVLVEVVASAVNRADLQRFYAPPPGASPPGLECSGRIA 70
Qy 71 KLGPNASTDLKVGDTGTFGFGVHGASQTDPKNGAFAEYARVYPLFYKSNLTHSTADEISEG 130
Db 71 ALPGVSV-GWAVGDEVNCLLAG-----GGVAEKVAV-----PAGQLLPV 108
Qy 131 PVK-NFESAASLP-VSLTT-AGVSLCHLHSGKMEWHPSTPQHTHPLINGGATAVGQQLI 187
Db 109 PEGIGLRQAALPEVTCVWSNVFMVAHL-----RPGET--LLVHGSSGIGTMAI 157
Qy 188 QVAKHINAYTKIVTVASKKHEKLLKSYGADVDYHDGAVIEQIKSYPNLQHVDAVGS 247
Db 158 QLAQAVGAKV-AVTAGTKEKLDFCAELGADVLVNYREQDFVEVR-----BATDGAGA 209
Qy 248 EDSIPEAYKYVTADSLPATLL 267
Db 210 D-----VLDNMGAKYL 221

RESULT 30
US-09-908-744-69
; Sequence 69 Application US/09908744
; Publication No. US20030068791A1
; GENERAL INFORMATION:
; APPLICANT: Misaenikov Andrei
; APPLICANT: Ojamo Heikki
; APPLICANT: Povelainen Mira
; APPLICANT: Gros H+kan
; APPLICANT: Toivari Mervi
; APPLICANT: Richard Peter
; APPLICANT: Ruohonen Laura
; APPLICANT: Koivuranta Kari
; APPLICANT: Londeborough John
; APPLICANT: Aristidou Aristos
; APPLICANT: Penttil, Merja
; APPLICANT: Plazanet-Menut Claire
; APPLICANT: Deutscher Josef
; TITLE OF INVENTION: Manufacture of Five-Carbon Sugars and Sugar Alcohols
; FILE REFERENCE: 1427.0010005
; CURRENT APPLICATION NUMBER: US/09/908,744
; CURRENT FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: PCT/FI01/00051
; PRIOR FILING DATE: 2001-01-22
; PRIOR APPLICATION NUMBER: US 09/488,581
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US 08/790,585
; PRIOR FILING DATE: 1997-01-29
; PRIOR APPLICATION NUMBER: US 08/368,395
; PRIOR FILING DATE: 1995-01-03

; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 11902
; LENGTH: 326
; TYPE: PRT
; ORGANISM: Bacillus halodurans
US-09-908-744-69

Query Match      7.5%; Score 144.5; DB 11; Length 352;
Best Local Similarity 24.4%; Pred. No. 5.9e-05;
Matches 76; Conservative 35; Mismatches 109; Indels 91; Gaps 14;

Qy 22 DYSVPELKEGTALVKV--EAVAGNPTDWKHIAVKIGPEGSILGCDIAGTVVKGPNASTD 79
Db 22 DLSVPEATDDKVLIKVAYTGICGSDIHTFKGEYKNPTTPVVLGHFSGQVVEVGANV-TK 80
Qy 80 LKVGDD-----TGFGFV-----HGASQTDPKNGAFAEYARVYPLFYKSN 118
Db 81 VKVGDRVTSETTF-YVCGECDYCKEYQNLCPHRKIGITQONGSMANY-----VLAREE 133
Qy 119 LTHSTADEISEGPKNFESAA-SLPVSLTTAGVSLCHLHSGKMEWHPSTPQHTHPLING 177
Db 134 SIHLLPDHLS-----YEGAMSEPLACCVHAMYQKSHLEK-----DTIIM 175
Qy 178 GATAVCQQLIOVAKHINAYTKIVTVASKGHE-KLLKSYGADVDYHD---AGVISOIKS 233
Db 176 GPGPIGLYLQIAKEIGAFVIMTGITKDAHRLALAKKLGADVIVDTMKEDLAKVNSBITD 235
Qy 234 KYPNLQHVDAVGSEDSIPEAYKVATDSLPAITLLEVVVPMTIESIPBEIRKDNVKIDITLL 293
Db 236 GY-GVDKVVDAAGA-----VPAVNASLP-----L 258
Qy 294 YRASGOEILLG 304
Db 259 IRKQGGFIQVG 269

Search completed: January 10, 2004, 10:16:57
Job time : 75 secs
```

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 10, 2004, 08:39:46 ; Search time 74 Seconds  
(without alignment)  
793.633 Million cell updates/sec

Title: US-10-081-644-2

Perfect score: 1920

Sequence: 1 MSVPTTQKAVIIIEGKAVK.....ALTEGIKEGKNVYVARL 370

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 90 summaries

Database : A Geneseq 19Jun03:\*

1: /SIDSI1/gcgcdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDSI1/gcgcdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDSI1/gcgcdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDSI1/gcgcdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDSI1/gcgcdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDSI1/gcgcdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDSI1/gcgcdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDSI1/gcgcdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDSI1/gcgcdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDSI1/gcgcdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDSI1/gcgcdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDSI1/gcgcdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDSI1/gcgcdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDSI1/gcgcdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDSI1/gcgcdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDSI1/gcgcdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDSI1/gcgcdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDSI1/gcgcdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDSI1/gcgcdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDSI1/gcgcdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDSI1/gcgcdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDSI1/gcgcdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDSI1/gcgcdata/geneseq/geneseq-emb1/AA2002.DAT.\*  
24: /SIDSI1/gcgcdata/geneseq/geneseq-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1055.5	55.5	376	23	ABG93187
2	473	24.6	361	23	ABG93300
3	310.5	16.2	131	23	ABP10927
4	266.5	13.9	363	21	AA196748
5	266	13.9	318	21	AA196748
6	266	13.9	318	21	AA196748
7	265	13.8	322	22	AAU35110
8	264	13.8	313	23	AB47714
9	261	13.6	386	21	AAG43467

10	254	13.2	309	21	AAG43468	Arabidopsis thalia
11	254	13.2	386	21	AAG22152	Arabidopsis thalia
12	247	12.9	309	21	AAG22153	Arabidopsis thalia
13	206	10.7	338	22	AAU35351	Enterococcus faeca
14	173	9.0	328	23	AB854030	Lactococcus lactis
15	171	8.9	421	23	ABU10941	Yeast selected int
16	168	8.8	337	22	AAG93071	C glutamicum prote
17	168	8.8	337	22	AAG79594	Corynebacterium gl
18	167	8.7	363	21	AAG28581	Arabidopsis thalia
19	167	8.7	363	21	AAG52443	Arabidopsis thalia
20	167	8.7	366	21	AAG28580	Arabidopsis thalia
21	167	8.7	366	21	AAG52442	Arabidopsis thalia
22	164.5	8.6	324	21	AAG39983	Arabidopsis thalia
23	164.5	8.6	329	21	AAG39982	Arabidopsis thalia
24	164.5	8.6	329	23	AB92914	Herbicidally activ
25	164.5	8.6	354	21	AAG39981	Arabidopsis thalia
26	162	8.4	262	21	AAG05068	Arabidopsis thalia
27	162	8.4	267	21	AAG05067	Arabidopsis thalia
28	162	8.4	292	21	AAG05066	Arabidopsis thalia
29	158.5	8.3	396	17	AAR97569	Interleukin-2 rece
30	158.5	8.3	413	22	AB118136	Human IL-2R-associ
31	155	8.1	117	21	AAB27050	Auxin-induced prom
32	155	8.1	117	23	AAU80747	Eucalyptus grandis
33	155	8.1	334	22	AAU59488	Propionibacterium
34	153.5	8.0	336	16	AAR71935	Sec. alcohol-dehyd
35	151.5	7.9	340	23	ABP38413	Staphylococcus epi
36	151	7.9	325	21	AAG17585	Arabidopsis thalia
37	151	7.9	325	21	AAG50141	Arabidopsis thalia
38	149.5	7.8	322	20	AA123394	Amino acid sequenc
39	149.5	7.8	332	21	AA159181	Human oxidoreducta
40	149	7.8	318	22	AAG89976	C glutamicum prote
41	148.5	7.7	348	21	AAG17994	Arabidopsis thalia
42	148.5	7.7	348	21	AAG27477	Arabidopsis thalia
43	148.5	7.7	348	23	AB92696	Herbicidally activ
44	148.5	7.7	370	21	AAG27476	Arabidopsis thalia
45	147.5	7.7	351	21	AAG05685	Arabidopsis thalia
46	147.5	7.7	362	21	AAG05684	Arabidopsis thalia
47	147	7.7	343	23	ABP28632	Arabidopsis thalia
48	145.5	7.6	357	23	ABP28623	Streptococcus poly
49	144.5	7.5	352	22	AAE05993	Enterococcus avium
50	144.5	7.5	403	22	AAU61595	Propionibacterium
51	143	7.4	332	23	AB48452	Listeria monocytog
52	142.5	7.4	348	23	AB806016	Pyrococcus horikos
53	142.5	7.4	367	23	AB806016	Monascus purpureus
54	141.5	7.4	337	13	AAR326874	Alcohol dehydrogen
55	141.5	7.4	337	14	AAR35044	Insertion element
56	141.5	7.4	348	22	AA896816	Putative P. abyssi
57	139.5	7.3	337	22	AAU36333	Pseudomonas aerugi
58	138.5	7.2	287	21	AAG28988	Arabidopsis thalia
59	138.5	7.2	291	21	AAG28987	Arabidopsis thalia
60	137.5	7.2	289	21	AAG20690	Arabidopsis thalia
61	137.5	7.2	308	21	AAG20689	Arabidopsis thalia
62	137.5	7.2	353	21	AAG20688	Arabidopsis thalia
63	137	7.1	308	21	AAG11238	Arabidopsis thalia
64	137	7.1	353	21	AAG11237	Arabidopsis thalia
65	137	7.1	356	21	AAG11236	Arabidopsis thalia
66	135.5	7.1	292	21	AAG05686	Arabidopsis thalia
67	134.5	7.0	306	22	AAG93028	C glutamicum prote
68	134.5	7.0	306	22	AA879226	Corynebacterium gl
69	133.5	7.0	336	21	AA195047	Candida albicans p
70	133	6.9	346	21	AAG04618	Arabidopsis thalia
71	132.5	6.9	396	17	AAR97570	Interleukin-2 rece
72	131.5	6.8	308	21	AAG47252	Arabidopsis thalia
73	131.5	6.8	333	21	AAG47251	Arabidopsis thalia
74	131.5	6.8	368	21	AAG47250	Arabidopsis thalia
75	131	6.8	333	22	AAU34133	Staphylococcus aur
76	131	6.8	335	22	AAU36558	Staphylococcus aur
77	131	6.8	363	20	AAW95109	Amino acid sequenc
78	131	6.8	377	22	AA19931	Human oxidoreducta
79	131	6.8	377	23	AA19931	Human Adhr-1 prote
80	130.5	6.8	340	23	AAE23267	Bacillus strain LN
81	130.5	6.8	377	23	ABP69791	Human polypeptide
82	129	6.7	334	23	ABU04705	Mycobacterium tube

83 129 6.7 343 22 AAU30322 Novel human secret  
84 128 6.7 292 21 AAG04620 Arabidopsis thalia  
85 128 6.7 307 21 AAG04619 Arabidopsis thalia  
86 127.5 6.6 362 22 ABB66217 Drosophila melanog  
87 127.5 6.6 362 22 ABB70568 Drosophila melanog  
88 127 6.6 338 23 ABP30052 Streptococcus poly  
89 126 6.6 2532 21 AAY96753 A. terreus LovF/Sc  
90 125.5 6.5 308 21 AAG47268 Arabidopsis thalia

ALIGNMENTS

RESULT 1  
ABG93187  
ID ABG93187 standard; Protein; 376 AA.  
XX AC ABG93187;  
XX XX  
DT 21-NOV-2002 (first entry)  
XX XX  
DE S. cerevisiae BAX-associated protein fragment SEQ ID 332.  
XX XX  
KW Bax; Bax-resistance; cytostatic; fungicide; immunosuppressive; virucide;  
KW vasotrophic; vaccine; gene therapy; proliferative disorder; cancer;  
KW apoptosis; fungal; yeast; infection; autoimmune disease; ischaemia;  
KW neurodegeneration; cell death.  
XX XX  
OS Saccharomyces cerevisiae.  
XX XX  
PN WO200264766-A2.  
XX XX  
PD 22-AUG-2002.  
XX XX  
PF 21-DEC-2001; 2001WO-EPI5398.  
XX XX  
PR 22-DEC-2000; 2000EP-0870318.  
XX XX  
PR 04-JAN-2001; 2001EP-0870002.  
XX XX  
PR 09-JAN-2001; 2001EP-0870003.  
XX XX  
PA (JANC ) JANSSEN PHARM NV.  
XX XX  
PI Contreras RH, Eberhardt I, Luyten WHML, Reekmans RJ;  
XX XX  
DR WPI; 2002-667002/71.  
XX XX  
DR N-PSDB; ABQ76453.  
XX XX  
PT New isolated nucleic acid representing a synthetic Bax gene, useful as  
PT medicament for treating, preventing and/or alleviating yeast or fungal  
PT infections or proliferative disorders, or for preventing apoptosis in  
PT certain diseases  
XX XX  
PS Claim 36; Figure 1; 344pp; English.  
XX XX  
CC This invention describes a novel nucleic acid representing a synthetic  
CC Bax gene. The Bax gene of the invention is useful for identifying  
CC Bax-resistant yeast or fungi, identifying, or obtaining and identifying  
CC Candida spp. sequences that are differentially expressed in a pathway  
CC eventually leading to programmed cell death or identifying inhibitors or  
CC inhibitor sequences of Bax-induced cell death. The products of the  
CC invention have cytostatic, fungicide; immunosuppressive, virucide and  
CC vasotrophic activity and can be used in vaccines or for gene therapy. The  
CC isolated nucleic acids, polypeptides, pharmaceutical compositions,  
CC antisense molecules and antibodies are useful as medicaments or in  
CC preparing a medicament for treating, preventing and/or alleviating  
CC diseases associated with yeast or fungi or proliferative disorders, such  
CC as cancer, or for preventing apoptosis in certain diseases. The compounds  
CC or polypeptides, or the genetically modified organism are useful for  
CC preparing a medicament for modifying the endogenic flora of humans and  
CC other mammals. The vaccine is useful for immunising against yeast or  
CC fungal infections. Apoptosis-related diseases include autoimmune disease,  
CC ischaemia, diseases related with viral infections or neurodegenerations.  
CC This sequence represents a polypeptide associated with the Bax gene

CC described in the disclosure of the invention.  
XX XX  
SQ Sequence 376 AA;  
Query Match 55.5%; Score 1065.5; DB 23; Length 376;  
Best Local Similarity 54.0%; Pred. No. 1.4e-88;  
Matches 201; Conservative 59; Mismatches 109; Indels 3; Gaps 2;  
QY 2 SVPTTOKAVIEGDKAVKTDYVPELKEGTALVKVEAVAGNPTDKHAYKIGPESIL 61  
DB 4 SIPETMKAVVIENGKAVVQDIPPELEBGFVLKTVAVAGNPTDKHIDFKIGPOGALL 63  
QY 62 GCDIAGTVVVKLGPN-STDCLKVGTGTFVHGASQTDPKNGAPAEYARVVPPLFYK--SN 118  
DB 64 GCDAAQIVKLGPNVDAARFAIGDIYGVHGSVFPSPNGAPAEYASISSETAYKPARE 123  
QY 119 LTHSTADEISEGPKVNFESAASLPVSLTTAGVSLCHHLGSKMEWHFSTTPQHTHPLIWGG 178  
DB 124 FRLCGKDKLPEGFVKSLEGAVALPVSLTTAGMILTHSFGLDMTWKPSKAQRDQPILPWGG 183  
QY 179 ATAVGQOLIQAQVAKHINAYTKIVTVASKKHEKLLKSYGADVDVHDHAGVIEQIKSKYPNL 238  
DB 184 ATAVGQMLQLAKLNGFSKIIIVASRKHKLKEYGADELFDYHDADVIEQIKKYNMI 243  
QY 239 QHVIDAVGSEDSIPEAYKVTADSLPATLLEVVPMTIESPEEIRKONKIDITLLYRASG 298  
DB 244 PVLVDCVSTETIQVYKCAADDLDATVVLTKEDIKEDRRQNVSEIGTLLYLG 303  
QY 299 QEILLGATRPASPEYHEATVFPKFINPHLNGDIHHNMIKVFSNGLDDVPALTEGIKE 358  
DB 304 NDVPFGTFTLPADPEYKEAAIKFIRKPNKINDEIHHIPVKVYKNGLDDIPQLLDDIKH 363  
QY 359 GKNKNVYVARL 370  
DB 364 GRNSGKLVAVL 375  
RESULT 2  
ABG93300  
ID ABG93300 standard; Protein; 361 AA.  
XX AC ABG93300;  
XX XX  
DT 21-NOV-2002 (first entry)  
XX XX  
DE C. albicans BAX-associated protein fragment SEQ ID 558.  
XX XX  
KW Bax; Bax-resistance; cytostatic; fungicide; immunosuppressive; virucide;  
KW vasotrophic; vaccine; gene therapy; proliferative disorder; cancer;  
KW apoptosis; fungal; yeast; infection; autoimmune disease; ischaemia;  
KW neurodegeneration; cell death.  
XX XX  
OS Candida albicans.  
XX XX  
PN WO200264766-A2.  
XX XX  
PD 22-AUG-2002.  
XX XX  
PF 21-DEC-2001; 2001WO-EPI5398.  
XX XX  
PR 22-DEC-2000; 2000EP-0870318.  
XX XX  
PR 04-JAN-2001; 2001EP-0870002.  
XX XX  
PR 09-JAN-2001; 2001EP-0870003.  
XX XX  
PA (JANC ) JANSSEN PHARM NV.  
XX XX  
PI Contreras RH, Eberhardt I, Luyten WHML, Reekmans RJ;  
XX XX  
DR WPI; 2002-667002/71.  
XX XX  
DR N-PSDB; ABQ76566.  
XX XX  
PT New isolated nucleic acid representing a synthetic Bax gene, useful as  
PT medicament for treating, preventing and/or alleviating yeast or fungal

PT infections or proliferative disorders, or for preventing apoptosis in  
PT certain diseases -  
PS Claim 36; Figure 2; 344pp; English.  
XX This invention describes a novel nucleic acid representing a synthetic  
CC Bax gene. The Bax gene of the invention is useful for identifying  
CC Bax-resistant yeast or fungi, identifying, or obtaining and identifying  
CC Candida spp. sequences that are differentially expressed in a pathway  
CC eventually leading to programmed cell death or identifying inhibitors or  
CC inhibitor sequences of Bax-induced cell death. The products of the  
CC invention have cytostatic, fungicide, immunosuppressive, virucide and  
CC vasotropic activity and can be used in vaccines or for gene therapy. The  
CC isolated nucleic acids, polypeptides, pharmaceutical compositions,  
CC antisense molecules and antibodies are useful as medicaments or in  
CC preparing a medicament for treating, preventing and/or alleviating  
CC diseases associated with yeast or fungi or proliferative disorders, such  
CC as cancer, or for preventing apoptosis in certain diseases. The compounds  
CC or polypeptides, or the genetically modified organism are useful for  
CC preparing a medicament for modifying the endogenous flora of humans and  
CC other mammals. The vaccine is useful for immunising against yeast or  
CC fungal infections. Apoptosis-related diseases include autoimmune disease,  
CC ischemia, diseases related with viral infections or neurodegenerations.  
CC This sequence represents a polypeptide associated with the Bax gene  
CC described in the disclosure of the invention.  
XX  
SQ Sequence 361 AA;

Query Match 24.6%; Score 473; DB 23; Length 361;

Best Local Similarity 34.1%; Pred. No. 1.6e-34;

Matches 128; Conservative 67; Mismatches 150; Indels 30; Gaps 14;

QY 8 KAVIIIEG---DKAVVKTDSVPELKEGTALVKVAVAGNPTDWKHIAVIGPEGSILG 62  
DB 2 KAAIISGPEPYQLAEIK-DIPOQKIKENILKAVAFINPTDWKHIVTQLSGPDVVG 60  
QY 63 CDIAGTVVVLGPNASTDLKVGDTGFGFVHGASQTDPKNGAFAYARYVP--PLFYKSNLT 120  
DB 61 CDVSGIIEEVGSQV-TGFAKGDVTSAPITG--NRSPTGAFAEVAVDPATSIKYNKFE 117  
QY 121 HSTADRISEGVKNFSAASLPVSLTTAGVSLCHLGSKMWEHPSTPQHTHPLINGGAT 180  
DB 118 HLTNLQVSE--IHSFEGASINLGLVTVGLSFHYLRIDNKKQPG-----DSLILWGAT 170  
QY 181 AVCQQLIQVAKHNAYT-KIVTVASKKHKLKSYGADVDYVDHAGVIFQIKSKYPNLQ 239  
DB 171 ATGVLAIQVAKV--YNLKVITASPKNHTLLQLGADYVDFDGDADVNNKKN-IGQIK 227  
QY 240 HVIDAVGSEDSIPEAVKVTADSLPATLLEVPVMTIESI-PEEIRKONVKIDITLLYRASG 298  
DB 228 FALDTIATPETFOKVYDSTEGSQEVFIDSLAGLDYRSIANDARGDQVHWGHTIACLASL 287  
QY 299 QEILLGATRPAPSEYHEATVKFKFVNPHLNGDIIHNMKIVFSNGLDVPALTEGIKE 358  
DB 288 KERTVFENYVQPELDDPTQWQKVQPID-ID-RIKHTNLKLLNEGLDSV---SEGLEL 343  
QY 359 GKKNV---KYVARL 370  
DB 344 SRNKLSAEKVPRV 358

RESULT 3

ABP10927

ID ABP10927 standard; Protein; 131 AA.

XX

AC

XX

DB

XX

XX

DE

XX

KW

KW

KW degenerative disorder; osteoarthritis; neurodegenerative disorder;  
KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;  
KW hypertension; hypothyroidism; cholesterol ester storage disease;  
KW immune deficiency; immune disorder; infectious disease;  
KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;  
KW myasthenia gravis.  
XX  
OS Homo sapiens.  
XX  
XX WO200192523-A2.  
XX  
PD 06-DEC-2001.  
XX  
XX 29-MAY-2001; 2001WO-US10836.  
XX  
XX 30-MAY-2000; 2000US-206132P.  
XX  
XX 29-AUG-2000; 2000US-228716P.  
XX  
XX (CURA-) CURAGEN CORP.  
XX  
XX Shimkets RA, Leach MD;  
XX  
XX WPI; 2002-106308/14.  
XX  
XX N-PSDB; ABN26679.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Query Match 16.2%; Score 310.5; DB 23; Length 131;

Best Local Similarity 49.2%; Pred. No. 2.5e-20;

Matches 64; Conservative 16; Mismatches 47; Indels 3; Gaps 2;

QY 74 PNA-STDLKVGDTGFGFVHGASQTDPKNGAFAYARYVPLFYK--SNLTHSTADEISEG 130

DB 2 PNVDAAFAIGDIYGVHIGASVRFPSNGAFAYSAISSETAYKPAEPFLCGDKLKEG 61

QY 131 PVKNFESAASLPVSLTTAGVSLCHLGSKMWEHPSTPQHTHPLINGGATAVGQQLIOVA 190

DB 62 PVKSLGAVSLPVSLTTAGMLTHSFGLDWTWPKAQDPILFWGGATAVGQMLQLA 121

QY 191 KHINAYTKIV 200

Db	122	KLNGFSKII 131	
AA996748			
AA996748		standard; Protein; 363 AA.	
XX			
AC	AA996748;		
XX			
XX			
DT	09-OCT-2000	(first entry)	
XX			
DE			
XX			
XX			
KW	Lovastatin; D4B segment; monacolin J; dehydrogenase; LovC; anti-lipemic;		
KW	HMG-CoA reductase inhibitor; anti-hypercholesterolaemic; anti-fungal.		
XX			
OS	Aspergillus terreus.		
XX			
XX			
PN	WO200037629-A2.		
XX			
PD	29-JUN-2000.		
XX			
PF	13-DEC-1999;	99WO-US29583.	
XX			
PR	18-DEC-1998;	98US-0215694.	
XX			
XX	(WISC ) WISCONSIN ALUMNI RES FOUND.		
XX			
PI	Hutchinson RC, Kennedy J, Park C;		
XX			
XX	WPI; 2000-442660/38.		
DR	N-PSDB; AAA51300.		
XX			
XX			
PT	Increasing lovastatin or monacolin J production in an organism, for use		
PT	as antihypercholesterolemic or antifungal agents, comprises		
PT	transforming the organism with a D4B segment		
XX			
PS	Disclosure; Page 62-63; 116pp; English.		
XX			
CC	The proteins shown in AA996744-60 are encoded by 17 genes from a		
CC	cluster in Aspergillus terreus (ATCC 20542), which flank the NPKS		
CC	(nonaketide polyketide synthase) gene, which is known to be required for		
CC	lovastatin production. The NPKS gene is contained within the context of		
CC	the entire gene cluster but is not indicated here (see US5744350). The		
CC	genes and proteins are named "ORF" or "Lov", where "Lov" signifies genes		
CC	shown to be essential for lovastatin production. The portion of the gene		
CC	cluster between ORF1 and the mid-region of LovF is referred to as the		
CC	"D4B segment". Increasing lovastatin, or monacolin J, production in a		
CC	lovastatin-producing organism, comprises transforming the organism with		
CC	a D4B segment, and expressing it (claimed). Lovastatin will also be		
CC	produced in non-lovastatin producing organisms (e.g. A. nidulans) by		
CC	transformation with the D4B segment and the entire LovF gene. The		
CC	methods are used to increase biosynthetic production of lovastatin (with		
CC	an at least 5-fold increase) which is an anti-hypercholesterolaemic		
CC	agent, and also has some anti-fungal activity. Lovastatin inhibits the		
CC	conversion of hydroxymethylglutarylcoenzyme A (HMG-CoA) into mevalonate		
CC	by HMG-CoA reductase. The methods can also be used to increase production		
CC	of monacolin J (claimed), which has anti-fungal activity.		
XX			
SQ	Sequence 363 AA;		
Query Match.	13.9%;	Score 266.5; DB 21; Length 363;	
Best Local Similarity	26.3%;	Pred. No. 1.1e-15;	
Matches 103;	Conservative 69;	Mismatches 156;	Indels 63; Gaps 17;
4	PTTQKAVII--EGDKAVVKTDSVPELKEGTALVKEAVAGNPTDKWHIAYKIGESILG	62	
9	PPQQTALTVNDHDEVTWNAAPCPMLPRDQVYRVEAVAINPSDTK-MRQQTATPWAFLG	67	
63	CDTAGTVKLGPNASTDLKVGDTGFGFVHGASQDTPKNGAFABY-----ARVYPLPYKSN	118	
68	TDYAGTVWVAGSDV-THIQVGDRVYGAQNMCPRTPDQGAFSQYTVTRGRWA-----	119	
119	LTHSTADEISEGFEVKNFESAASLPVSLTTAGVSLCHHLGSKMEWHPEST---POHTHP--L	173	
120	-----KIPKG--LSFEQAALPAGISTAGLAM-KLGLPLP-SPSADQPPTHSKPVVY	168	
174	LIWGGATAVGQQLIQVAKHINAYTKIVTVASKKHEKLLKSYGADDVFDYHDAGVISQIKS	233	
169	LVYGGSTATATVTMQLR-LSGYIPIAT-CSPHNFOLAKSRGAEEVFDYRAPNLAQTIRT	226	
234	-KYPNLQHVIDAVGSEDSIPEAYKYVTADSLPATLLEVVPMPTIESIPEE-IRKDNVKIDIT	291	
227	YTKNNLRYALDCITNVESTTFCFAAIGRAGGHY-----VSLNPPFEHAATROMVTTDWT	280	
292	LLYRASQGBILLGATRF-----PASPEVHEATVKFVKFINPHLNNGDIHHMNI	339	
281	-----LGPITFGEGSTWPAPYGRPGSEERQFGEDLWRIAGQLVEDGRLVHPL	329	
340	KVFSNGLDDVPALTEGIEKGNKNVYVARL	370	
330	RVVQGGFDHIKQGMELVRKGLSGELVVRL	360	
RESULT 5			
AA996748			
ID	AA996748	standard; Protein; 318 AA.	
XX			
AC	AA996748;		
XX			
DT	02-FEB-2001	(first entry)	
XX			
DE			
XX			
KW	Auxin-induced promoter protein #1.		
XX			
KW	Promoter; eucalyptus; pine; gene transcription.		
XX			
OS	Eucalyptus grandis.		
XX			
PN	WO200058474-A1.		
XX			
PD	05-OCT-2000.		
XX			
PF	24-FEB-2000;	2000WO-NZ00018.	
XX			
PR	25-MAR-1999;	99US-0276599.	
PR	30-JUL-1999;	99US-0146591.	
XX			
PA	(GENE-) GENESIS RES & DEV CORP LTD.		
PA	(FLET-) FLETCHER CHALLENGE FORESTS LTD.		
XX			
PI	Perera R, Rice SJ, Eagleton CK;		
XX			
DR	WPI; 2000-647236/62.		
DR	N-PSDB; AAC62778.		
XX			
XX			
PT	Novel promoter sequences useful for modulating transcription of plant		
PT	DNA sequences of interest and production of polypeptides		
XX			
PS	Claim 4; Pages 64-65; 93pp; English.		
XX			
CC	The present invention relates to promoter sequences from eucalyptus and		
CC	pine. The present protein is encoded by one such promoter. The coding		
CC	sequence for the present protein is useful for modulating the		
CC	transcription of DNA sequences of interest. The sequences may also be		
CC	used to tag or identify an organism or its reproductive material.		
XX			
SQ	Sequence 318 AA;		
Query Match	13.9%;	Score 266; DB 21; Length 318;	
Best Local Similarity	33.6%;	Pred. No. 1e-15;	
Matches 91;	Conservative 31;	Mismatches 93;	Indels 56; Gaps 11;
3	VPTQKAVII--EGDKAVV---KTDVSVPELKEGTALVKEAVAGNPTDKWHIAYKIGPE	57	
7	IPTKMKAVYRHBGVDVAVNLGLDPELKEPQGVLVKVLAAALNPIDTARVKGVIKLP	66	







PR	27-AUG-1999;	99US-0151080.	PR	25-FEB-1999;	99US-0121825.
PR	30-AUG-1999;	99US-0151303.	PR	05-MAR-1999;	99US-0123180.
PR	31-AUG-1999;	99US-0151438.	PR	09-MAR-1999;	99US-0123548.
PR	01-SEP-1999;	99US-0151930.	PR	23-MAR-1999;	99US-0125788.
PR	07-SEP-1999;	99US-0152363.	PR	25-MAR-1999;	99US-0126264.
PR	10-SEP-1999;	99US-0153070.	PR	29-MAR-1999;	99US-0126785.
PR	13-SEP-1999;	99US-0153758.	PR	01-APR-1999;	99US-0127462.
PR	15-SEP-1999;	99US-0154018.	PR	06-APR-1999;	99US-0128234.
PR	16-SEP-1999;	99US-0154039.	PR	08-APR-1999;	99US-0128714.
PR	20-SEP-1999;	99US-0154779.	PR	16-APR-1999;	99US-0129845.
PR	22-SEP-1999;	99US-0155139.	PR	19-APR-1999;	99US-0130077.
PR	23-SEP-1999;	99US-0155486.	PR	21-APR-1999;	99US-0130449.
PR	24-SEP-1999;	99US-0155659.	PR	23-APR-1999;	99US-0130510.
PR	28-SEP-1999;	99US-0156458.	PR	23-APR-1999;	99US-0130891.
PR	29-SEP-1999;	99US-0156596.	PR	28-APR-1999;	99US-0131449.
PR	04-OCT-1999;	99US-0157117.	PR	30-APR-1999;	99US-0132048.
PR	05-OCT-1999;	99US-0157753.	PR	04-MAY-1999;	99US-0132407.
PR	06-OCT-1999;	99US-0157865.	PR	05-MAY-1999;	99US-0132485.
PR	07-OCT-1999;	99US-0158029.	PR	06-MAY-1999;	99US-0132486.
PR	08-OCT-1999;	99US-0158232.	PR	07-MAY-1999;	99US-0132487.
PR	12-OCT-1999;	99US-0158369.	PR	11-MAY-1999;	99US-0132863.
PR	13-OCT-1999;	99US-0159233.	PR	14-MAY-1999;	99US-0132866.
PR	13-OCT-1999;	99US-0159294.	PR	14-MAY-1999;	99US-0134218.
PR	13-OCT-1999;	99US-0159295.	PR	14-MAY-1999;	99US-0134219.
PR	14-OCT-1999;	99US-0159329.	PR	14-MAY-1999;	99US-0134221.
PR	14-OCT-1999;	99US-0159330.	PR	14-MAY-1999;	99US-0134370.
PR	14-OCT-1999;	99US-0159331.	PR	18-MAY-1999;	99US-0134768.
PR	14-OCT-1999;	99US-0159637.	PR	19-MAY-1999;	99US-0134941.
PR	14-OCT-1999;	99US-0159638.	PR	20-MAY-1999;	99US-0135124.
PR	18-OCT-1999;	99US-0159584.	PR	21-MAY-1999;	99US-0135353.
PR	21-OCT-1999;	99US-0160741.	PR	24-MAY-1999;	99US-0135629.
PR	21-OCT-1999;	99US-0160767.	PR	25-MAY-1999;	99US-0136021.
PR	21-OCT-1999;	99US-0160768.	PR	27-MAY-1999;	99US-0136392.
PR	21-OCT-1999;	99US-0160770.	PR	28-MAY-1999;	99US-0136782.
PR	21-OCT-1999;	99US-0160814.	PR	01-JUN-1999;	99US-0137222.
PR	21-OCT-1999;	99US-0160815.	PR	03-JUN-1999;	99US-0137528.
PR	22-OCT-1999;	99US-0160980.	PR	04-JUN-1999;	99US-0137502.
PR	22-OCT-1999;	99US-0160981.	PR	07-JUN-1999;	99US-0137724.
PR	22-OCT-1999;	99US-0160989.	PR	08-JUN-1999;	99US-0138094.
PR	23-OCT-1999;	99US-0161404.	PR	10-JUN-1999;	99US-0138540.
PR	25-OCT-1999;	99US-0161405.	PR	14-JUN-1999;	99US-0138847.
PR	25-OCT-1999;	99US-0161406.	PR	16-JUN-1999;	99US-0139119.
PR	26-OCT-1999;	99US-0161359.	PR	17-JUN-1999;	99US-0139452.
PR	26-OCT-1999;	99US-0161360.	PR	18-JUN-1999;	99US-0139454.
PR	28-OCT-1999;	99US-0161361.	PR	18-JUN-1999;	99US-0139455.
PR	28-OCT-1999;	99US-0161920.	PR	18-JUN-1999;	99US-0139456.
PR	28-OCT-1999;	99US-0161992.			
PR	28-OCT-1999;	99US-0161993.			
PR	28-OCT-1999;	99US-0162142.			
Query Match 13.68; Score 261; DB 21; Length 386;			RESULT 10		
Best Local Similarity 32.28; Pred. No. 3.9e-15;			AAG43468		
Matches 88; Conservative 37; Mismatches 104; Indels 44; Gaps 11;			ID AAG43468 standard; Protein; 309 AA.		
QY	2 SVPTOKAVTIEG----	DKAVKTDVSPVPELKEGTALVKVEAVAGNPTDVKHAYKI---	XX	AC AAG43468;	
Db	73 SIPKEMKAWYSYGGVDVLKESNVVPEIKEDQVLKVVAAALNPFVDKRRQKPKAT 132		XX	DT 18-OCT-2000 (first entry)	
QY	55 -GPGSILGCDIAGTVKVLGNASTDLKVGDTGFGFVHGASQTDPKN-GAFAYARVYPP 112		DE	Arabidopsis thaliana protein fragment SEQ ID NO: 54335.	
Db	133 DPLPTVPDVGAVGVKVG-SAVKDLKEGDEVVAVNVEKALEGPKQFGSLAEYTAEEK 191		XX	Protein identification; signal transduction pathway; metabolic pathway;	
QY	113 L--FYKSNLTHSTADEISEGPNFSAASLPVSLTTAGVSLCHHLSGKMEWHPSTPQHT 170		KW	hybridisation assay; genetic mapping; gene expression control; promoter;	
Db	192 LLALKPKNI-----DFAQAAGLPLAIETADEGLV-----RTEFSAG----- 227		KW	termination sequence.	
QY	171 HPLLWGATVAGQOLIQVAKHNAYTKIVTVASKKHEKLLKSGYADDVFDYHDAGVIEQ 230		XX	Arabidopsis thaliana.	
Db	228 KSLVLNAGGVGSLVLQIAKHVYGVASKVAATASTEKLELVRLSGADLAIDYTKEN-IED 286		OS	EP1033405-A2.	
QY	231 IKSYPNLQHVDAVGSSEDS-----IPEAYKVT 259		PD	06-SEP-2000.	
Db	287 LPDKY---DVVFDAGMCDKAVKVIKEGGKVA 316		XX	25-FEB-2000; 2000EP-0301439.	



```
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139464.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145921.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.

PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 22-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 13.2%; Score 254; DB 21; Length 309;
Best Local Similarity 32.8%; Pred. No. 1.2e-14;
Matches 84; Conservative 35; Mismatches 97; Indels 40; Gaps 10;

QY 15 DKAVVTDVSVPELKEGTALVKVEAVAGNPTDWHYAYKI----GPEGSTLGCDIAGTVV 70
Db 13 DVLKLESNIWPEIKEDQVLKIVVAAALNPVDARQKFKATDSPLTPVPGYDVAGVV 72
QY 71 KLGNASTDLKVGDTGFGFVHGASQTDPKN-GAFAYARVY PPL---FYKSNLTHSTADEI 127
Db 73 KVG-SAVKDLKEGDEVYANVSEKALEGPKQFGSLAETAVEEKLALKPKNI----- 123
QY 128 SEGPKNFESAASLPVSLTTAGVSLCHILGSKMEHPSTPQHTPLHLLWGATAVGOOLI 187
```

Db 124 -----DFAQAAGPLAETADEGLV-----RTEFSAG-----KSIILVINGAGGVGSLVI 167  
Qy 188 QVAKHINAYTKIVTVASKKHEKLLKSYGADVDYHDAGVIEQIKSKYPNLQHVDAVGS 247  
168 QLAHVYTGASKVATASTEKLELVRSIGADLAIDYTKEN-IEDLPDKY----DVVFDAIGM 223  
Qy 248 EDS-----IPEAYKVTA 259  
Db 224 CDKAVKVIKGGKWA 239

RESULT 11  
AAG22152  
ID AAG22152 standard; Protein; 386 AA.  
AC AAG22152;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 24972.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EF1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132484.  
PR 06-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 18-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144632.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 03-AUG-1999; 99US-0146389.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.



PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136332.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139432.  
PR 18-JUN-1999; 99US-0139434.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142330.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143352.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 28-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.

PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 14-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.

```
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 12.98; Score 247; DB 21; Length 309;
Best Local Similarity 32.08; Pred. No. 5.4e-14;
Matches 8; Conservative 35; Mismatches 99; Indels 40; Gaps 10;

QY 15 DKAVKTDVSPVPELKEGTALVKVEAVAGNPTDMKHIAKYI---GPGSILGCDIAGTVV 70
DB 13 DVLKLESINAIPEIKEDQVLKVVAGLNPVDKCRQKPKATDSPLTPVPGYDVAGVVV 72
QY 71 KLGNASTDLKVGDTGFGFVHGASQTDPKN-GAPAEYARVYPP--FYKSNLTHSTADEI 127
DB 73 KVG-SAVKDFKEGDEVYANSEKALEGPKQFGSLAEYTAVEEKLALKPKNI----- 123
QY 128 SEGPKVKNFESAASI.PVSLTTAGVSLCHHLGSKMEWHPSTPQHTPLLIWGATAVGQOLI 187
DB 124 -----DFAQAAGLPLAETADEGLV-----RTEFSAG-----KSILVLNGAGGVGSLMI 167
QY 188 QVAKHINAYTKIVTVASKHEKLKLSYGADDVDFYHDAGVIEQIKSKYPNLQHVDAVGS 247
DB 168 QLAHVYGVASKVAATASTGKLELVRSIGADLAIDYTKEN-IEDLPDKY---DVVFDAGM 223
QY 248 EDS-----IPEAYKVTA 259
DB 224 CDRAVKVIKEGGKVA 239

RESULT 13
AAU35351
ID AAU35351 standard; Protein; 338 AA.
AC AAU35351;
XX
XX
DT 14-FEB-2002 (first entry)
DE Enterococcus faecalis cellular proliferation protein #638.
KW Antisense; prokaryotic cellular proliferation protein;
KW antibiotic; antibacterial; drug design.
XX
XX OS Enterococcus faecalis.
XX WO200170955-A2.
XX
XX 27-SEP-2001.
XX
XX 21-MAR-2001; 2001WO-US09180.
XX
XX 21-MAR-2000; 2000US-191078P.
XX 23-MAY-2000; 2000US-206848P.
XX 26-MAY-2000; 2000US-207727P.
XX 23-OCT-2000; 2000US-242578P.
XX 27-NOV-2000; 2000US-253625P.
XX 22-DEC-2000; 2000US-257931P.
XX 16-FEB-2001; 2001US-269308P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
XX Yamamoto RT, Xu HH;
XX WPI; 2001-611495/70.
XX N-PSDB; AAS53210.
XX
XX New polynucleotides for the identification and development of
XX antibiotics, comprise sequences of antisense nucleic acids -
XX
XX Example 3; Seq ID No 10944; Sllpp; English.
XX
XX The invention relates to antisense inhibitors of genes essential to
XX prokaryotic cellular proliferation, their use in identifying the
```

```
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 338 AA;
SQ

Query Match 10.7%; Score 206; DB 22; Length 338;
Best Local Similarity 28.1%; Pred. No. 3.4e-10;
Matches 83; Conservative 39; Mismatches 87; Indels 86; Gaps 16;

QY 7 QKAVIIIEGDKAVVKTDSVPPELKEGTALVKVEAVAGNPTDMKH-----IAYKIGPEG 58
DB 12 QKELAIIE-----EVPLTIHNDVLVKIIAASINPIDLTKDGKVKMLLVQM---P 60
QY 59 SILGCDIAGTVVKGPNASTDLKVGDTGFGFVHGASQTDPKN--GAPAEYARV--YPPLF 114
DB 61 LILGSDFAGIWSVGKKVQ-NFRLGDAVYGRV-----PKNRVGTFAEYIAVDQAAM 112
QY 115 YKSNLTHSTADEISEGPKVKNFESAASLPVSLTTAGVSLCHHLGSKMEWHPSTPQHTPL 174
DB 113 KPKNLT-----FEERAAIPL-----VGLTSQALHDIMNVQPG-----QKVL 149
QY 175 IMGATAVGQOLIQAQKHINAYTKIVTVASKHEKLKLSYGADDVDFYHDAGVIEQIKSK 234
DB 150 IQAGSGGIGTIAIQLAKLAGY--VATTTSSKNKWEVQALGADEVIDYRTQN-FEEVLSD 206
QY 235 YPNLQHVDAVGSEDSIPEAYKVTAADSLPATLIE-----VVP-----MTIESIEEE 280
DB 207 Y---DYVFDTMG-----GTILEKAFSVVVKPGKVVTLSGIPNE 241

RESULT 14
ABB54030
ID ABB54030 standard; Protein; 328 AA.
XX
XX ABB54030;
XX
XX 16-MAY-2002 (first entry)
XX Lactococcus lactis protein qor.
XX
XX Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.
XX
XX Lactococcus lactis IL1403.
XX
XX FR2807446-A1.
XX
XX 12-OCT-2001.
XX
XX 11-APR-2000; 2000FR-0004630.
XX
XX 11-APR-2000; 2000FR-0004630.
XX (INRG ) INRA INST NAT RECH AGRONOMIQUE.
XX
XX Polotine A, Sorokine A, Renault P, Ehrlich SD;
XX WPI; 2002-043418/06.
XX
```

PT New nucleotide sequence useful in the identification or Lactococcus  
 PT lactis and related species -  
 PS Claim 6; SEQ ID No 732; 2504pp; French.  
 XX  
 CC The present invention is related to a Lactococcus lactis nucleotide  
 CC sequence (ABA90521) and related proteins (ABB53300-ABB55621). The  
 CC nucleic acid sequence is useful in the detection and/or amplification of  
 CC nucleic acid sequence, particularly to identify Lactococcus lactis or  
 CC related species. The proteins of the invention are useful for the  
 CC biosynthesis or biodegradation of a composition of interest. The  
 CC invention helps research in lactic bacteria, particularly useful in the  
 CC production of yogurt and cheese.  
 CC Note: The sequence data for this patent is based on equivalent patent  
 CC WO2001/77334 (published 18-OCT-2001) which is available in electronic  
 CC format directly from WIPO at ftp.wipo.int/pub/published\_pat\_sequences.  
 XX  
 SQ Sequence 328 AA;  
 Query Match 9.0%; Score 173; DB 23; Length 328;  
 Best Local Similarity 28.8%; Pred. No. 3.3e-07;  
 Matches 72; Conservative 43; Mismatches 91; Indels 44; Gaps 11;  
 QY 8 KAV-IIEGDKAVKTDVSPPELKEGTALVKVEAVAGNFTDWHIAYKIGPEGS-----IL 61  
 DB 2 KAIGTIDSKNFIDFEIEKPLRPHDLLIKVEAISINPVDTK---VRKGIKGNLAEPKIL 58  
 QY 62 GCDIAGTVVKLGNASTDLKVGDTGP--GFVHGASQDTPKNGFAFAEVARVYPL--FYKS 117  
 DB 59 GWDGLGTVVELGSETKL-FKVGDKVFWAGDV-----TRSGNAEFQAVDERIVGFAPK 110  
 QY 118 NLTHSTADEISEGPKNFESAASLPVSLTTAGVSLCHLHSGKMEWHPSTPQHTHPLLLMG 177  
 DB 111 NLTK-----EKAVMLPSTLTAAYELLPEKL-----EVTDKSKGKSLIIN 150  
 QY 178 GATAVGQQLIQVAKHINAYTKIVTVASKKHEKLKSYGADDVFDYHDAGV--IEQIKSKY 235  
 DB 151 GAGGVGSVAIQAKNA-GLTVIATASNPAIEWVKNGFADYTVNHHKLVPPQVHELGFKF 209  
 QY 236 PNQHVUIDAV 245  
 DB 210 VDYLILNAV 219  
 RESULT 15  
 ABJ10941  
 ID ABJ10941 standard; Protein; 421 AA.  
 AC ABJ10941;  
 XX  
 DT 10-DEC-2002 (first entry)  
 XX  
 DE Yeast selected interacting domain protein SEQ ID NO: 16.  
 XX  
 KW Yeast; protein-protein interaction; Selected Interacting Domain;  
 KW STD (RTM); secretion yield; cancer; neurodegenerative disease; fungicide;  
 KW cytotstatic; neuroprotective.  
 XX  
 OS Saccharomyces cerevisiae.  
 XX  
 PN WO200266504-A2.  
 XX  
 PD 29-AUG-2002.  
 XX  
 PF 14-FEB-2002; 2002WO-EP02299.  
 XX  
 PR 16-FEB-2001; 2001US-269266P.  
 XX  
 PA (HYBR-) HYBRIGENICS.  
 XX  
 PI Legrain P;  
 XX  
 DR WPI; 2002-674913/72.

DR N-PSDB; ABT11258.  
 XX  
 PT New protein-protein complexes of Saccharomyces cerevisiae, useful in  
 PT drug screening or development, for developing yeast strains with better  
 PT secretion yield of protein, or in gene therapy (e.g. to treat Candida  
 PT infection or cancer) -  
 XX  
 PS Claim 6; Page 41; 357pp; English.  
 XX  
 CC The present invention relates to complexes between Saccharomyces  
 CC cerevisiae Selected Interacting Domain (SID (RTM)) proteins and coding  
 CC sequences. The protein complexes of S. cerevisiae are useful in drug  
 CC development, in screening drugs or agents that modulate the interaction  
 CC of proteins, for developing yeast strains with better secretion yield of  
 CC protein, and in gene therapy. The protein complexes, polypeptides and  
 CC polynucleotides are useful for preventing or treating Candida infection,  
 CC cancer or neurodegenerative diseases in humans or animals. The present  
 CC sequence is a protein of the invention.  
 XX  
 SQ Sequence 421 AA;  
 Query Match 8.9%; Score 171; DB 23; Length 421;  
 Best Local Similarity 46.5%; Pred. No. 7.3e-07;  
 Matches 33; Conservative 12; Mismatches 26; Indels 0; Gaps 0;  
 QY 300 EILLGATREFPASPEYHEATVKFKFNPILNNGDIHMMIKVFSNGLDDVPALTEGKEG 359  
 DB 2 EVFPGNITUPADSEAKAAIKFIKIPNKINDQIRHIPRVVYKNGLCDVPHILDKIKYG 61  
 QY 360 KKNKVKYVARL 370  
 DB 62 KNSGEKLVAVL 72  
 RESULT 16  
 AAG93071  
 ID AAG93071 standard; Protein; 337 AA.  
 XX  
 AC AAG93071;  
 XX  
 DT 26-SEP-2001 (first entry)  
 XX  
 DE C glutamicum protein fragment SEQ ID NO: 6825.  
 XX  
 KW Corynebacterium; amino acid synthesis; vitamin; saccharide;  
 KW organic acid synthesis.  
 XX  
 OS Corynebacterium glutamicum.  
 XX  
 PN EP1108790-A2.  
 XX  
 PD 20-JUN-2001.  
 XX  
 PF 18-DEC-2000; 2000EP-0127688.  
 XX  
 PR 16-DEC-1999; 99JP-0377484.  
 PR 07-APR-2000; 2000JP-0159162.  
 PR 03-AUG-2000; 2000JP-0280988.  
 XX  
 PA (KYOW ) KYOWA HAKKO KOGYO KK.  
 XX  
 PI Nakagawa S, Mizoguchi H, Ando S, Hayaashi M, Ochiai K, Yokoi H;  
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;  
 XX  
 DR WPI; 2001-376931/40.  
 DR N-PSDB; AAH68290.  
 XX  
 PT Novel polynucleotides derived from Coryneform bacteria, for identifying  
 PT mutation point of a gene, measuring expression of a gene, analysing  
 PT expression profile or pattern of a gene and identifying homologous gene  
 PT -  
 XX  
 PS Claim 17; SEQ ID NO: 6825; 246pp + Sequence Listing; English.









PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139494.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139889.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140921.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142300.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145152.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.

PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0156659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 08-OCT-1999; 99US-0157865.  
PR 08-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 18-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 26-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 8.7%; Score 167; DB 21; Length 363;  
Best Local Similarity 26.1%; Pred. No. 1.4e-06;

Matches	72;	Conservative	47;	Mismatches	91;	Indels	66;	Gaps	15
---------	-----	--------------	-----	------------	-----	--------	-----	------	----

  

QY 2 SVPTTKAVAIIE--GDKAV--VKTDVSPPELKEGTALVKVEAVAGNPDMK----- 48  
| | | | | : : : : :  
Db 24 SVFTGCRAVILPRFGGFEVFELRENVFPVNLPNNEVLVKAKAVSNPLDCRIRAGYGRSV 83  
| | | | | : : : : :  
QY 49 ---HIAYKIPEGSGILGCDTACTGVKLGPNASTDLKVGDTCGFQFVHGASQTDPKNAPAE 105  
| | | | | : : : : :  
Db 84 FQPHLPI-----IVGRDVSGEVAATGTSVKSL-KVGQEVFGALHPTA----LRGTVD 131  
| | | | | : : : : :  
QY 106 YARVYPPLFYKSNLTHSTADRISEGP--VKNFESAASLPVSLTTAGVSLCHHLGSKMEWH 163  
| | | | | : : : : :  
Db 132 YGIL-----SEDELTKPSSISHVE-ASAIPFAALTAWRAL----- 166  
| | | | | : : : : :  
QY 164 PSTPQHT--HPLLIWGATAVGOOLIOVAKHINAYTKIVTVASKKHKKLLKSYGADDVFD 221  
| | | | | : : : : :  
Db 167 KSNARITEGQRLLVFGGGGAVGPSAILQLAVASGCHVTASCVCVGTK-DRLAA-GAEQAVD 224  
| | | | | : : : : :  
QY 222 YHDAGVTEQKSQKVPNLQHVDAVGSEDSDIPEAYKV 257  
| | | | | : : : : :  
Db 225 YTTFEDIELAVKGKF---DAVLDTTIGG---PETERI 253  
| | | | | : : : : :

RESULT 20  
AAG28580  
ID AAG28580 standard; Protein; 366 AA.  
XX AC AAG28580;  
XX XX  
DT 17-OCT-2000 (first entry)  
XX XX  
XX Arabidopsis thaliana protein fragment SEQ ID NO: 33852.  
XX Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
XX termination sequence.  
XX XX  
OS Arabidopsis thaliana.  
XX XX  
PN EP1033405-A2.  
XX XX  
PD 06-SEP-2000.  
XX XX  
XX 25-FEB-2000; 2000EP-0301439.  
XX XX  
PR 25-FEB-1999; 99US-0121825.  
PR 03-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.

RESULT 21  
AAG52442

PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140981.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144335.  
PR 19-JUL-1999; 99US-0144335.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145087.

PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 13-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 28-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.

```
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 28-OCT-1999; 99US-0162142.
PR 28-OCT-1999; 99US-0162142.

Query Match      8.7%; Score 167; DB 21; Length 366;
Best Local Similarity 26.1%; Pred. No. 1.4e-06;
Matches 72; Conservative 47; Mismatches 91; Indels 66; Gaps 15;

QY 2 SVPTTQKAVIE--GDKAV--VKTDVSVPELKEGTALVKVEAVAGNPTDMK----- 48
Db 27 SVFTGCAVILPRFGGEVFEFLRENVPVNLNPNNEVLVKAKAVSNPLDCRIRAGYGRSV 86
QY 49 ---HIAYKIGPEGILGCDIAGTVVKLGPNASTDVKVGDTCGFVHGASQTDPKNGAFAE 105
Db 87 FQPHLP1-----IVGRDVSGEVAAIGTSVKSLKVGQEVFGALHPTA----LRGTYTD 134
QY 106 YARYVPLFYKSNLTHSTADBSISGP--VKNFESAASLPVSLTTPAGVSLCHILGSKNMEWH 163
Db 135 YGIL-----SEDLTEKPSISHSVE--ASAIPFAALTAMRAL----- 169
QY 164 PSTQHT--HPLLWGGATAGQOLIQAOKHINAYTKIVTVASKKHEKLKSKYGADVD 221
Db 170 KSNARITEGQRLLVFGGGAVGFSAIQIYASGCHVTASCVGQTK-DRILAA-GAEQAVD 227
QY 222 YHDAGVTEQIKSKYPNLQHVLDVAGSEDSIPEAYKV 257
Db 228 YTTEDIELAVKQKP---DAVLDTIGG-----PETERI 256

RESULT 22
AAG39983
ID AAG39983 standard; Protein; 324 AA.
AC AAG39983;
XX
XX
DT 18-OCT-2000 (first entry)
XX
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 49547.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
XX Arabidopsis thaliana.
XX
XX EPI033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0131825.
XX 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 05-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 23-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 08-JUN-1999; 99US-0138094.
PR 08-JUN-1999; 99US-0137724.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 18-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
```

PR 19-JUL-1999;	99US-0144335.	PR 14-OCT-1999;	99US-0159330.
PR 20-JUL-1999;	99US-0144352.	PR 14-OCT-1999;	99US-0159331.
PR 20-JUL-1999;	99US-0144632.	PR 14-OCT-1999;	99US-0159637.
PR 20-JUL-1999;	99US-0144884.	PR 14-OCT-1999;	99US-0159638.
PR 21-JUL-1999;	99US-0144814.	PR 18-OCT-1999;	99US-0159584.
PR 21-JUL-1999;	99US-0145086.	PR 21-OCT-1999;	99US-0160741.
PR 21-JUL-1999;	99US-0145088.	PR 21-OCT-1999;	99US-0160767.
PR 21-JUL-1999;	99US-0145085.	PR 21-OCT-1999;	99US-0160768.
PR 22-JUL-1999;	99US-0145087.	PR 21-OCT-1999;	99US-0160770.
PR 22-JUL-1999;	99US-0145192.	PR 21-OCT-1999;	99US-0160814.
PR 23-JUL-1999;	99US-0145145.	PR 21-OCT-1999;	99US-0160815.
PR 23-JUL-1999;	99US-0145218.	PR 22-OCT-1999;	99US-0160980.
PR 23-JUL-1999;	99US-0145224.	PR 22-OCT-1999;	99US-0160981.
PR 26-JUL-1999;	99US-0145276.	PR 22-OCT-1999;	99US-0160989.
PR 27-JUL-1999;	99US-0145913.	PR 25-OCT-1999;	99US-0161404.
PR 27-JUL-1999;	99US-0145918.	PR 25-OCT-1999;	99US-0161405.
PR 27-JUL-1999;	99US-0145919.	PR 25-OCT-1999;	99US-0161406.
PR 28-JUL-1999;	99US-0145951.	PR 26-OCT-1999;	99US-0161359.
PR 02-AUG-1999;	99US-0146386.	PR 26-OCT-1999;	99US-0161360.
PR 02-AUG-1999;	99US-0146388.	PR 26-OCT-1999;	99US-0161361.
PR 02-AUG-1999;	99US-0146389.	PR 28-OCT-1999;	99US-0161920.
PR 03-AUG-1999;	99US-0147038.	PR 28-OCT-1999;	99US-0161992.
PR 04-AUG-1999;	99US-0147204.	PR 28-OCT-1999;	99US-0161993.
PR 04-AUG-1999;	99US-0147302.	PR 29-OCT-1999;	99US-0162142.
PR 05-AUG-1999;	99US-0147192.	Query Match 8.6%; Score 164.5; DB 21; Length 324;	
PR 05-AUG-1999;	99US-0147260.	Best Local Similarity 23.4%; Pred. No. 1.9e-06;	
PR 06-AUG-1999;	99US-0147303.	Matches 80; Conservative 56; Mismatches 141; Indels 65; Gaps 16;	
PR 06-AUG-1999;	99US-0147416.		
PR 09-AUG-1999;	99US-0147493.		
PR 09-AUG-1999;	99US-0147935.		
PR 10-AUG-1999;	99US-0148171.		
PR 11-AUG-1999;	99US-0148319.		
PR 12-AUG-1999;	99US-0148341.		
PR 13-AUG-1999;	99US-0148565.		
PR 13-AUG-1999;	99US-0148684.		
PR 16-AUG-1999;	99US-0149368.		
PR 17-AUG-1999;	99US-0149175.		
PR 18-AUG-1999;	99US-0149426.		
PR 20-AUG-1999;	99US-0149722.		
PR 20-AUG-1999;	99US-0149723.		
PR 20-AUG-1999;	99US-0149929.		
PR 23-AUG-1999;	99US-0149902.		
PR 23-AUG-1999;	99US-0149930.		
PR 25-AUG-1999;	99US-0150566.		
PR 26-AUG-1999;	99US-0150884.		
PR 27-AUG-1999;	99US-0151065.		
PR 27-AUG-1999;	99US-0151066.		
PR 27-AUG-1999;	99US-0151080.		
PR 30-AUG-1999;	99US-0151303.		
PR 31-AUG-1999;	99US-0151438.		
PR 01-SEP-1999;	99US-0151930.		
PR 07-SEP-1999;	99US-0152363.		
PR 10-SEP-1999;	99US-0153070.		
PR 13-SEP-1999;	99US-0153758.		
PR 15-SEP-1999;	99US-0154018.		
PR 16-SEP-1999;	99US-0154039.		
PR 20-SEP-1999;	99US-0154779.		
PR 22-SEP-1999;	99US-0155139.		
PR 23-SEP-1999;	99US-0155486.		
PR 24-SEP-1999;	99US-0155659.		
PR 28-SEP-1999;	99US-0156458.		
PR 29-SEP-1999;	99US-0156596.		
PR 04-OCT-1999;	99US-0157117.		
PR 05-OCT-1999;	99US-0157753.		
PR 06-OCT-1999;	99US-0157865.		
PR 07-OCT-1999;	99US-0158029.		
PR 08-OCT-1999;	99US-0158232.		
PR 12-OCT-1999;	99US-0158369.		
PR 13-OCT-1999;	99US-0159293.		
PR 13-OCT-1999;	99US-0159294.		
PR 13-OCT-1999;	99US-0159295.		
PR 14-OCT-1999;	99US-0159329.		

Qy 14	GDKAVVKTDSVPELKEGTALVKVEAVAGNPTDKHIAIKGP-----EGSILGCDIAGT 68
Db 12	GAAGLEHVQVPVTPPKSNVCLKLEATSLNPVDWKIQKGMIRPFLPRKFCIPATDVAGE 71
Qy 69	VVKLGPNASTDLKVGDTGFGFVHGASQTDPKNGAFAEYARVYPLFYKSNLTHSTADEIS 128
Db 72	VVEVSGVK-NPKAGDKVAVLSHLG-----GGGLAEFA-----VATEKLTVRQEVG 119
Qy 129	EGPVKNFESAASLPVSLTTAGVSLCHHLGSKMEWHPTQHTPLLIWGATAVGOOLIQ 188
Db 120	AA-----EAAALPVAGLTALQALTNPAGLKLD--GTGKAN-ILVTAASGGVGHYAVQ 169
Qy 189	VAKHINAYKIVTVASKGHEKLLKSYGADDDVFIY---HDAGVIEQIKSKY----- 235
Db 170	LAKLANAH--VTATCGARNIEFVKSLGADDEVLDYKTPGGAALKSPSGKKYDAVVCANGI 227
Qy 236	-----PNLQH---VIDAVGSEDSIPEAYKVTADSLPATLLEVVPMPTIESIPPEIRKQNV 286
Db 228	PFSVFEPNLSENGKVIDITPGPNAM---WTYAVKKTMSKKQLVPLLL--IP---KAENL 279
Qy 287	KIDITLLYRASGOEILLGATRPPASPEYHEATVKFVKFINPH 328
Db 280	EFMVLVKEGKVKTVI--DSKHPLS-----KAEDAWAKSIDGH 315

RESULT 23

AAG39982

ID AAG39982 standard; Protein; 329 AA.

XX AAG39982;

XX 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 49546.

XX Protein identification; signal transduction pathway; metabolic pathway;

KW hybridisation assay; genetic mapping; gene expression control; promoter;

XX termination sequence.

OS Arabidopsis thaliana.

FN EP1033405-A2.

PD 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.  
PF 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136302.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140323.  
PR 23-JUN-1999; 99US-0140334.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149502.  
PR 23-AUG-1999; 99US-0149930.  
PR 23-AUG-1999; 99US-0150566.  
PR 25-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 30-AUG-1999; 99US-0151080.  
PR 31-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.



```

PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match      8.6%; Score 164.5; DB 21; Length 329;
Best Local Similarity 23.4%; Pred. No. 2e-06;
Matches 80; Conservative 56; Mismatches 141; Indels 65; Gaps 16;

Qy 14 GDKAVKTDVSVPELKEGTALVKVEAVAGNPTDKHIAKYIGP-----EGSILGCDIAGT 68
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
17 GAAGLEHVQVPVPTPKSNEVCLKEATSLNPVDWKIQGMRPFLPRKFCIPATDVAGE 76
Qy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
69 VVKLGNASTDLKVGDTGFGFVHGASQTDPKNGAFAYARYVPPFLFYKSNLTHSTADEIS 128
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
77 VVEGSGVK-NFKAGDKVAVLSHLG-----GGGLAEFA-----VATEKLTVKRPQEVG 124
Qy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
129 EGPVKNPESAASLPVSLTTAGVSLCHLGSKMWHPSPTPOHTPLLIWGATAVGOQLIQ 188
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
125 AA-----EAAALPVAGLTALQALTNPAGLKLD---GTGKKAN-ILVTAASGGVGHYAVQ 174
Qy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
189 VAKHINAYTKIVTVASKKHEKLLKSGADDDVFDY---HDAGVIEQISKY-----235
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
175 LAKLANAH--VTATCGARNIEPVKSLGADVDLYKTPEGAALKSPSGKKYDAVVHCANGI 232
Qy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
236 -----PNLQH---VIDAVGSEDSIPEAYKVTADSLPATLLEVVPMPTIESIPERKDNV 286
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
287 KIDITLLYRASGOEILLGATRPASPPEYHEATVKFVKFINPH 328
285 EFMVNLVKEGKVKTVI--DSKHPLS-----KAEDAWAKSIDGH 320

```

RESULT 24

ABB92914

ID ABB92914 standard; Protein; 329 AA.

XX ABB92914;

AC ABB92914;

XX

```

DT 31-MAY-2002 (first entry)
XX
DE Herbicidally active polypeptide SEQ ID NO 2125.
XX
KW Herbicidal; plant; agriculture; herbicide.
XX
OS Arabidopsis thaliana.
XX
PN WO200210210-A2.
XX
PD 07-FEB-2002.
XX
PF 28-AUG-2001; 2001WO-EP09892.
XX
PR 28-AUG-2001; 2001WO-EP09892.
XX
PA (FARB ) BAYER AG.
XX
PI Tietjen K, Weidler M;
XX
DR WPI; 2002-269010/31.
XX
PT Identifying plant target proteins for herbicidally active compounds, comprising aligning and comparing nucleic acid or amino acid sequences from plant with nucleic acid or amino acid sequences from non-plant organisms -
XX
PS Claim 5; SEQ ID NO 2125; 261pp + Sequence Listing; English.
XX
CC The invention relates to identifying target proteins (ABB90790-ABB94016) for herbicidally active compounds, comprising aligning and comparing nucleic acid or amino acid sequences from plant with nucleic acid or amino acid sequences from non-plant organisms using suitable search parameters, where plant sequences having an E-value greater by a factor of 3 than the E-value of most similar non-plant sequences are selected. The polypeptides or nucleic acids encoding them are useful for identifying modulators. The identified modulators are useful as herbicides.
XX
SQ Sequence 329 AA;

Query Match      8.6%; Score 164.5; DB 23; Length 329;
Best Local Similarity 23.4%; Pred. No. 2e-06;
Matches 80; Conservative 56; Mismatches 141; Indels 65; Gaps 16;

Qy 14 GDKAVKTDVSVPELKEGTALVKVEAVAGNPTDKHIAKYIGP-----EGSILGCDIAGT 68
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
17 GAAGLEHVQVPVPTPKSNEVCLKEATSLNPVDWKIQGMRPFLPRKFCIPATDVAGE 76
Qy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
69 VVKLGNASTDLKVGDTGFGFVHGASQTDPKNGAFAYARYVPPFLFYKSNLTHSTADEIS 128
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
77 VVEGSGVK-NFKAGDKVAVLSHLG-----GGGLAEFA-----VATEKLTVKRPQEVG 124
Qy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
129 EGPVKNPESAASLPVSLTTAGVSLCHLGSKMWHPSPTPOHTPLLIWGATAVGOQLIQ 188
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
125 AA-----EAAALPVAGLTALQALTNPAGLKLD---GTGKKAN-ILVTAASGGVGHYAVQ 174
Qy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
189 VAKHINAYTKIVTVASKKHEKLLKSGADDDVFDY---HDAGVIEQISKY-----235
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
175 LAKLANAH--VTATCGARNIEPVKSLGADVDLYKTPEGAALKSPSGKKYDAVVHCANGI 232
Qy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
236 -----PNLQH---VIDAVGSEDSIPEAYKVTADSLPATLLEVVPMPTIESIPERKDNV 286
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
287 KIDITLLYRASGOEILLGATRPASPPEYHEATVKFVKFINPH 328
285 EFMVNLVKEGKVKTVI--DSKHPLS-----KAEDAWAKSIDGH 320

RESULT 25
AAG39981
ID AAG39981 standard; Protein; 354 AA.

```

XX AAG39981; 18-JUN-1999; 99US-0139462.  
AC 18-JUN-1999; 99US-0139463.  
XX 18-JUN-1999; 99US-0139750.  
DT 18-JUN-1999; 99US-0139763.  
XX 21-JUN-1999; 99US-0139817.  
XX 22-JUN-1999; 99US-0139899.  
XX 23-JUN-1999; 99US-0140353.  
XX 23-JUN-1999; 99US-0140354.  
XX 24-JUN-1999; 99US-0140695.  
XX 28-JUN-1999; 99US-0140823.  
OS 29-JUN-1999; 99US-0140991.  
XX 30-JUN-1999; 99US-0141287.  
PN 01-JUL-1999; 99US-0141842.  
XX 01-JUL-1999; 99US-0142154.  
XX 02-JUL-1999; 99US-0142055.  
XX 06-JUL-1999; 99US-0142390.  
XX 08-JUL-1999; 99US-0142803.  
XX 09-JUL-1999; 99US-0142920.  
XX 12-JUL-1999; 99US-0142977.  
XX 13-JUL-1999; 99US-0143542.  
XX 14-JUL-1999; 99US-0143624.  
XX 15-JUL-1999; 99US-0144005.  
XX 16-JUL-1999; 99US-0144085.  
XX 16-JUL-1999; 99US-0144086.  
XX 19-JUL-1999; 99US-0144325.  
XX 19-JUL-1999; 99US-0144331.  
XX 19-JUL-1999; 99US-0144332.  
XX 19-JUL-1999; 99US-0144333.  
XX 19-JUL-1999; 99US-0144334.  
XX 20-JUL-1999; 99US-0144335.  
XX 20-JUL-1999; 99US-0144352.  
XX 20-JUL-1999; 99US-0144632.  
XX 20-JUL-1999; 99US-0144884.  
XX 21-JUL-1999; 99US-0144814.  
XX 21-JUL-1999; 99US-0145086.  
XX 21-JUL-1999; 99US-0145088.  
XX 22-JUL-1999; 99US-0145085.  
XX 22-JUL-1999; 99US-0145087.  
XX 22-JUL-1999; 99US-0145089.  
XX 22-JUL-1999; 99US-0145192.  
XX 23-JUL-1999; 99US-0145145.  
XX 23-JUL-1999; 99US-0145218.  
XX 23-JUL-1999; 99US-0145224.  
XX 26-JUL-1999; 99US-0145276.  
XX 27-JUL-1999; 99US-0145913.  
XX 27-JUL-1999; 99US-0145918.  
XX 27-JUL-1999; 99US-0145919.  
XX 28-JUL-1999; 99US-0145951.  
XX 02-AUG-1999; 99US-0146386.  
XX 02-AUG-1999; 99US-0146388.  
XX 02-AUG-1999; 99US-0146389.  
XX 03-AUG-1999; 99US-0147038.  
XX 04-AUG-1999; 99US-0147204.  
XX 04-AUG-1999; 99US-0147302.  
XX 05-AUG-1999; 99US-0147192.  
XX 05-AUG-1999; 99US-0147260.  
XX 06-AUG-1999; 99US-0147303.  
XX 06-AUG-1999; 99US-0147416.  
XX 09-AUG-1999; 99US-0147493.  
XX 10-AUG-1999; 99US-0147935.  
XX 10-AUG-1999; 99US-0148171.  
XX 11-AUG-1999; 99US-0148319.  
XX 12-AUG-1999; 99US-0148341.  
XX 13-AUG-1999; 99US-0148565.  
XX 13-AUG-1999; 99US-0148684.  
XX 16-AUG-1999; 99US-0149368.  
XX 17-AUG-1999; 99US-0149175.  
XX 18-AUG-1999; 99US-0149426.  
XX 20-AUG-1999; 99US-0149722.  
XX 20-AUG-1999; 99US-0149723.  
XX 20-AUG-1999; 99US-0149929.  
XX 23-AUG-1999; 99US-0149902.  
XX 23-AUG-1999; 99US-0149930.

XX Arabidopsis thaliana protein fragment SEQ ID NO: 49545.  
XX Protein identification; signal transduction pathway; metabolic pathway;  
XX hybridisation assay; genetic mapping; gene expression control; promoter;  
XX termination sequence.  
XX Arabidopsis thaliana.  
XX EPI033405-A2.  
XX 06-SEP-2000.  
XX 25-FEB-2000; 2000EP-0301439.  
XX 25-FEB-1999; 99US-0121825.  
XX 05-MAR-1999; 99US-0123180.  
XX 09-MAR-1999; 99US-0123548.  
XX 23-MAR-1999; 99US-0125788.  
XX 25-MAR-1999; 99US-0126264.  
XX 29-MAR-1999; 99US-0126785.  
XX 01-APR-1999; 99US-0127462.  
XX 06-APR-1999; 99US-0128234.  
XX 08-APR-1999; 99US-0128714.  
XX 16-APR-1999; 99US-0129845.  
XX 19-APR-1999; 99US-0130077.  
XX 21-APR-1999; 99US-0130449.  
XX 23-APR-1999; 99US-0130510.  
XX 23-APR-1999; 99US-0130891.  
XX 28-APR-1999; 99US-0131449.  
XX 30-APR-1999; 99US-0132048.  
XX 30-APR-1999; 99US-0132407.  
XX 04-MAY-1999; 99US-0132484.  
XX 05-MAY-1999; 99US-0132485.  
XX 06-MAY-1999; 99US-0132486.  
XX 06-MAY-1999; 99US-0132487.  
XX 07-MAY-1999; 99US-0132863.  
XX 11-MAY-1999; 99US-0134256.  
XX 14-MAY-1999; 99US-0134218.  
XX 14-MAY-1999; 99US-0134219.  
XX 14-MAY-1999; 99US-0134221.  
XX 14-MAY-1999; 99US-0134370.  
XX 18-MAY-1999; 99US-0134768.  
XX 19-MAY-1999; 99US-0134941.  
XX 20-MAY-1999; 99US-0135124.  
XX 21-MAY-1999; 99US-0135353.  
XX 24-MAY-1999; 99US-0135629.  
XX 25-MAY-1999; 99US-0136021.  
XX 27-MAY-1999; 99US-0136392.  
XX 28-MAY-1999; 99US-0136782.  
XX 01-JUN-1999; 99US-0137222.  
XX 03-JUN-1999; 99US-0137528.  
XX 04-JUN-1999; 99US-0137502.  
XX 07-JUN-1999; 99US-0137724.  
XX 08-JUN-1999; 99US-0138094.  
XX 10-JUN-1999; 99US-0138540.  
XX 10-JUN-1999; 99US-0138847.  
XX 14-JUN-1999; 99US-0139119.  
XX 16-JUN-1999; 99US-0139452.  
XX 16-JUN-1999; 99US-0139453.  
XX 17-JUN-1999; 99US-0139492.  
XX 18-JUN-1999; 99US-0139454.  
XX 18-JUN-1999; 99US-0139455.  
XX 18-JUN-1999; 99US-0139456.  
XX 18-JUN-1999; 99US-0139457.  
XX 18-JUN-1999; 99US-0139458.  
XX 18-JUN-1999; 99US-0139459.  
XX 18-JUN-1999; 99US-0139460.  
XX 18-JUN-1999; 99US-0139461.

```
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 06-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match      8.6%; Score 164.5; DB 21; Length 354;
Best Local Similarity 23.4%; Pred. No. 2.2e-06;
Matches 80; Conservative 56; Mismatches 141; Indels 65; Gaps 16;

QY 14 GDKAVKTDVSVPLKSGTALVKVEAVAGNPTDWKHIAKIGP-----EGSILGCDIAGT 68
DB 42 GAAGLEHVQVPVPTPKSNEVCLKEATSLNPVDWKIQKGMIRPFLPRKFPFCIPATDVAGE 101
QY 69 VVKLGPNASTDLKVGDTGFGFVHGASQTDPKNGAFAYARYVPLPFYKSNLTHSTADEIS 128
DB 102 VVEVGSVK-NFRKAGDKVAVLSHLG-----GGGLAEFA-----VATEKLTVRPOEVG 149
QY 129 EGPVKNFESAASLPSVLTAGVSIKCHLGSXMEWHPTPQHTPLLIWGGATAVGQQLIQ 188
DB 150 AA-----EAAALPVAGLTALQALTNPAGLKLD---GTGKKAN-ILVTAASGGVGHYAVQ 199
QY 189 VAKHINAYTKIVTVASKKHEKLLKSYGADVDYD---HDAGVIBQIKSKY-----235
DB 200 LAKLANAH--VTATCGARNTEFVKSLGADELDYKTPTEGAALKSPGKYDAVVHCANGI 257

236 -----PNLQH-----VIDAVGSEDSIPEAYKVTADSLPATLLEVVPMTIETSPPEIRKDNV 286
258 PFSVPFNPNSKGVKIDITPGPNAM---WTYAVKKITMSKKQLVPLLL--IP---KAENL 309
QY 287 KIDITILYRASGOEILLGATRFPASPPEYHEATVKFVKFINPH 328
DB 310 EFMVNLVKEGKVKVTI--DSKHPLS----KAEDAWAKSIDGH 345

RESULT 26
AAG05068
ID AAG05068 standard; Protein; 262 AA.
XX
AC AAG05068;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 1312.
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
```



Db	12	GAAGLEHVQVPVPTPKNEVCLKLEATSLNPVDWIKIQKGMIRPFLPRKPCIPATDVAGE	71
Qy	69	VVKLGPNASTDLKVGDTGFGFVHGASQTDPRKNGAFAYRYVPELFYKSNLTHSTADEIS	128
Db	72	VVQVGSGVK-NFKAGDKVAVLSHLG-----GGGLAEFA-----VATEKLTVKRPQEVG	119
Qy	129	EGPVKNTFESAASLPSVLTTRAGVSILCHLGSQKNEWHPSTPQTHPLLWGGTAVGQQLIQ	188
Db	120	AA-----EAAALPVAGLTALQALTNPAGLKLD---GTGKKAN-ILVTAASGGVGHVAVQ	169
Qy	189	VAKHINAYTKIVTVASKKHEKLLSGYADDVFDY---HDAGVIEQIKSKYPENLOHVIDAV	245
Db	170	LAKLANAH--VTATCGARNIEFVKSGLGADEVLDYKTPBGAALKSPSGKKYDAVVHVCANGI	227
RESULT 27			
AAG05067	ID	AAG05067 standard; Protein; 267 AA.	
XX	AC	AAG05067;	
XX	DT	17-OCT-2000 (first entry)	
XX	XX	Arabidopsis thaliana protein fragment SEQ ID NO: 1331.	
DE	XX	Protein identification; signal transduction pathway; metabolic pathway;	
KW	KW	hybridisation assay; genetic mapping; gene expression control; promoter;	
KW	KW	termination sequence.	
XX	XX	Arabidopsis thaliana.	
OS	XX	EP1033405-A2.	
PN	XX	06-SEP-2000.	
PD	XX	25-FEB-2000; 2000EP-0301439.	
PF	XX	25-FEB-1999; 99US-0121825.	
PR	PR	05-MAR-1999; 99US-0123180.	
PR	PR	09-MAR-1999; 99US-0123548.	
PR	PR	23-MAR-1999; 99US-0125788.	
PR	PR	29-MAR-1999; 99US-0126264.	
PR	PR	01-APR-1999; 99US-0126785.	
PR	PR	06-APR-1999; 99US-0128234.	
PR	PR	08-APR-1999; 99US-0128714.	
PR	PR	16-APR-1999; 99US-0129845.	
PR	PR	19-APR-1999; 99US-0130077.	
PR	PR	21-APR-1999; 99US-0130449.	
PR	PR	23-APR-1999; 99US-0130510.	
PR	PR	23-APR-1999; 99US-0130891.	
PR	PR	30-APR-1999; 99US-0131449.	
PR	PR	30-APR-1999; 99US-0132048.	
PR	PR	30-APR-1999; 99US-0132407.	
PR	PR	04-MAY-1999; 99US-0132484.	
PR	PR	05-MAY-1999; 99US-0132485.	
PR	PR	06-MAY-1999; 99US-0132486.	
PR	PR	06-MAY-1999; 99US-0132487.	
PR	PR	07-MAY-1999; 99US-0132863.	
PR	PR	11-MAY-1999; 99US-0134256.	
PR	PR	14-MAY-1999; 99US-0134218.	
PR	PR	14-MAY-1999; 99US-0134219.	
PR	PR	14-MAY-1999; 99US-0134221.	
PR	PR	14-MAY-1999; 99US-0134370.	
PR	PR	18-MAY-1999; 99US-0134766.	
PR	PR	19-MAY-1999; 99US-0134941.	
PR	PR	20-MAY-1999; 99US-0135124.	
PR	PR	21-MAY-1999; 99US-0135353.	
PR	PR	24-MAY-1999; 99US-0135629.	
PR	PR	25-MAY-1999; 99US-0136021.	
PR	PR	27-MAY-1999; 99US-0136392.	
PR	PR	28-MAY-1999; 99US-0136782.	
PR	PR	01-JUN-1999; 99US-0137222.	
PR	PR	05-AUG-1999; 99US-0147192.	
PR	PR	05-AUG-1999; 99US-0147302.	
PR	PR	04-AUG-1999; 99US-0147204.	
PR	PR	03-AUG-1999; 99US-0146389.	
PR	PR	02-AUG-1999; 99US-0146388.	
PR	PR	02-AUG-1999; 99US-0145913.	
PR	PR	27-JUL-1999; 99US-0145918.	
PR	PR	28-JUL-1999; 99US-0145951.	
PR	PR	27-JUL-1999; 99US-0145919.	
PR	PR	27-JUL-1999; 99US-0145918.	
PR	PR	22-JUL-1999; 99US-0145192.	
PR	PR	23-JUL-1999; 99US-0145145.	
PR	PR	23-JUL-1999; 99US-0145218.	
PR	PR	23-JUL-1999; 99US-0145224.	
PR	PR	26-JUL-1999; 99US-0145276.	
PR	PR	27-JUL-1999; 99US-0145913.	
PR	PR	27-JUL-1999; 99US-0145918.	
PR	PR	27-JUL-1999; 99US-0145919.	
PR	PR	28-JUL-1999; 99US-0145951.	
PR	PR	02-AUG-1999; 99US-0146388.	
PR	PR	02-AUG-1999; 99US-0146389.	
PR	PR	03-AUG-1999; 99US-0146389.	
PR	PR	04-AUG-1999; 99US-0147036.	
PR	PR	04-AUG-1999; 99US-0147204.	
PR	PR	05-AUG-1999; 99US-0147192.	
PR	PR	05-AUG-1999; 99US-0147302.	
PR	PR	05-AUG-1999; 99US-0147192.	
PR	PR	05-AUG-1999; 99US-0147302.	
PR	PR	05-AUG-1999; 99US-0147192.	
PR	PR	05-AUG-1999; 99US-0147302.	
PR	PR	05-AUG-1999; 99US-0147192.	
PR	PR	05-AUG-1999; 99US-0147302.	
PR	PR	05-AUG-1999; 99US-0147192.	



PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145085.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 16-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 18-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.





antibodies against the polypeptides, methods of detecting the nucleotides or polypeptides in a sample, and methods of identifying compounds which bind to polypeptides of the invention. Although novel, many of the polypeptides of the invention have homology to known proteins, thereby giving an insight into their probable biological activities, and hence potential therapeutic applications. The polypeptides of the invention may have various activities, including cytokine, cell proliferation or cell differentiation activities; stem cell growth factor activity; haematopoietic regulatory activity; tissue growth factor activity; immunomodulatory activity; activin- or inhibin-related activities; chemotactic or chemokinetic activities; haemostatic, thrombotic or thrombolytic activities; receptor or ligand activities; or may be involved in oncogenesis, cancer cell proliferation or metastasis. Depending on their biological activities, polypeptides and nucleotides of the invention are useful for preventing, treating or ameliorating medical conditions, e.g., by protein or gene therapy. Such conditions include cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell disorders), chronic inflammatory conditions (e.g., asthma or arthritis), proliferative retinopathy, atherosclerosis, coronary heart disease, arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal vascular growth. Polypeptides involved with tissue regeneration and repair (or nucleic acids encoding them) may be used to promote wound healing (e.g., of burns, incisions and ulcers), while those with immunomodulatory activities may be used in the treatment of viral, bacterial and fungal infections in addition to immune disorders. Polypeptides with growth factor activity may be used in cell cultures to promote cell growth. For example, such polypeptides may be used to manipulate stem cells in culture to give rise to neuroepithelial cells that can be used to augment or replace cells damaged by illness, autoimmune disease or accidental damage. The polypeptides and nucleotides may also be used in the diagnosis of the above conditions, and in drug screening techniques. The present sequence represents a novel human polypeptide of the invention.

Sequence 413 AA;

Query Match		8.3%; Score 158.5; DB 22; Length 413;
Best Local Similarity		27.8%; Pred. No. 9.8e-06;
Matches		71; Conservative 39; Mismatches 96; Indels 49; Gaps 12;
Qy	34 LKVEAVAGNPTDWK-----	-----HIYKIGPEGSI-LGCDIAGTVVKLGN 75
Db	92 IVKHAASVNPIDVNMRSYGATALNMKRDPLHYKIK-GEFPLTLGRDVSQVVMCEGLD 150	
Qy	76 ASTDLKVGDTGFGFVHGASQTDPKNGAPAEYARVYVPLFYKSNLTHSTADEISSEGPVK-N 134	
Db	151 VKY-FKPGDEWV-----	AAVPPWQGLSEFVVV-----SGNEVSHKPKSLT 191
Qy	135 FESAASLPVSLTTAGVSLCHHLGSKMEWHPSTPQTHPLLIWGGATAVGQQLIQVAKHIN 194	
Db	192 HTQAASLPYVALTAW-SAINKVGGLNDKNCCTGKR----	VLILGASGGVGTFAIQVMKAWD 246
Qy	195 AYTQIVTVASKHEKLLKSGYADDVFDHAGVIEQIKSKYPNLOHVDAVG---SEDSIP 252	
Db	247 AH--VTAVCSQDASELVKRLGADDVIDYKSGSVEEQLSKLP-PDFILDNVGGSTETWAP 303	
Qy	253 EAYKVTADSLPATLL 267	
Db	304 DFLKWSGATVTLV 318	

Search completed: January 10, 2004, 10:04:54  
Job time : 80 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 10, 2004, 09:56:13 ; Search time 37 Seconds  
(without alignments)  
961.687 Million cell updates/sec

Title: US-10-081-644-2

Perfect score: 1920

Sequence: 1 MSVPTTQKAVIEGKAVVK.....ALTEGKGNKNKYVARL 370

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 90 summaries

Database :

PIR\_76.\*

1: PIR1.\*

2: PIR2.\*

3: PIR3.\*

4: PIR4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1065.5	55.5	376	2 S55149	hypothetical prote
2	1021.5	53.2	376	2 S59418	hypothetical prote
3	1015.5	52.9	368	2 S19414	hypothetical prote
4	264	13.8	313	2 AE1151	oxidoreductase hom
5	260	13.5	313	2 AF1510	auxin-induced prot
6	259	13.5	317	2 T10824	quinone oxidoreduc
7	254	13.2	309	2 E86371	zinc-binding dehyd
8	218	11.4	353	2 AF3199	quinone oxidoreduc
9	208	10.8	324	2 C83695	quinone oxidoreduc
10	193	10.1	343	1 C70418	probable alcohol d
11	187.5	9.8	322	2 G83766	quinone oxidoreduc
12	187	9.7	334	2 AH3214	zinc-binding oxido
13	176	9.2	331	2 B90457	alcohol dehydrogen
14	175.5	9.1	348	2 T18210	alcohol dehydrogen
15	173.5	9.0	328	2 AG3182	zinc-binding dehyd
16	173	9.0	328	2 E86714	quinone oxidoreduc
17	172.5	9.0	339	2 B93742	alginatase lyase BH0
18	171.5	8.9	338	2 AD2699	zinc-binding dehyd
19	171.5	8.9	338	2 D97481	probable oxidoredu
20	170.5	8.9	336	2 AH3289	NADPH2:quinone red
21	164.5	8.6	329	2 T10203	hypothetical prote
22	164.5	8.6	348	2 F97459	probable quinone o
23	164	8.5	337	2 A82309	quinone oxidoreduc
24	162	8.4	327	2 A12677	quinone oxidoreduc
25	161.5	8.4	328	2 F70871	probable quinone o
26	161	8.4	329	1 FN0448	zeta-crystallin /
27	159.5	8.3	318	2 AC1858	hypothetical prote
28	159	8.3	329	1 CYGP2	zeta-crystallin /
29	158.5	8.3	331	1 A54932	zeta-crystallin /

#### RESULT 1

S55149

N;Hypothetical protein YNL134c - yeast (Saccharomyces cerevisiae)

C;Species: Saccharomyces cerevisiae

C;Date: 08-Jul-1995 #sequence,revision 01-Sep-1995 #text\_change 19-Apr-2002

C;Accession: S55149; S59254; S63079

#### ALIGNMENTS

30	157	8.2	328	2	T21985	hypothetical prote
31	156.5	8.2	8563	2	T30226	polyketide synthas
32	156.5	8.2	10223	2	T30225	polyketide synthas
33	156	8.1	328	2	E70695	probable oxidoredu
34	154	8.0	365	2	S50409	hypothetical prote
35	153.5	8.0	326	2	T36504	probable quinone o
36	153.5	8.0	345	2	S57611	probable NADPH2:qu
37	153.5	8.0	348	1	S32521	alcohol dehydrogen
38	152	7.9	336	2	AH0201	probable zinc-bind
39	151	7.9	325	2	T40264	zinc-binding dehyd
40	151	7.9	325	2	T05166	quinone reductase
41	149.5	7.8	388	2	C75441	probable NADPH qui
42	148.5	7.7	348	2	T49047	quinone reductase-
43	147	7.7	328	2	E83504	probable oxidoredu
44	147	7.7	447	2	T35960	crotonyl-CoA reduc
45	146.5	7.6	338	2	H90403	alcohol dehydrogen
46	146.5	7.6	342	2	S57612	probable NADPH2:qu
47	146.5	7.6	351	2	G86389	probable allyl alc
48	146	7.6	343	2	T51554	quinone oxidoreduc
49	145.5	7.6	350	2	S09475	alcohol dehydrogen
50	144.5	7.5	6260	2	T30228	polyketide synthas
51	143.5	7.5	335	2	AB2174	hypothetical prote
52	143.5	7.5	335	2	E90014	hypothetical prote
53	143.5	7.5	350	1	S52153	alcohol dehydrogen
54	143	7.4	332	2	AE1396	zinc-binding dehyd
55	142.5	7.4	348	2	H71110	probable dehydroge
56	141.5	7.4	337	1	A42654	alcohol dehydrogen
57	141.5	7.4	348	2	G75049	L-threonine 3-dehy
58	141	7.3	341	2	G86037	threonine dehydrog
59	141	7.3	371	2	AB0407	probable zinc-bind
60	140	7.3	341	1	DEECTH	L-threonine 3-dehy
61	140	7.3	379	2	JN0013	synaptic vesicle m
62	139.5	7.3	337	2	G83200	probable oxidoredu
63	139.5	7.3	346	2	T51551	quinone oxidoreduc
64	139	7.2	341	2	P91190	threonine dehydrog
65	138.5	7.2	325	2	G87344	alcohol dehydrogen
66	137.5	7.2	326	2	F95888	probable NADPH2:qu
67	137.5	7.2	349	2	B90285	alcohol dehydrogen
68	137	7.1	359	2	H95892	probable oxidoredu
69	136	7.1	332	2	AH1771	zinc-binding dehyd
70	136	7.1	358	2	T51555	quinone oxidoreduc
71	134.5	7.0	374	1	S17252	alcohol dehydrogen
72	134	7.0	321	2	E87715	quinone oxidoreduc
73	134	7.0	336	2	F86923	probable oxidoredu
74	134	7.0	411	2	H95851	probable alcohol d
75	132	6.9	334	1	S45904	probable NADPH2:qu
76	131.5	6.8	342	2	AD3391	alcohol dehydrogen
77	131.5	6.8	375	1	S54458	alcohol dehydrogen
78	131	6.8	380	2	E84238	quinone oxidoreduc
79	131	6.8	447	2	S72400	trans-2-enoyl-CoA
80	131	6.8	4151	2	G70944	probable polyketid
81	130.5	6.8	348	1	DEBYA2	alcohol dehydrogen
82	130.5	6.8	353	1	A45052	L-iditol 2-dehydro
83	130	6.8	348	1	S32484	L-iditol 2-dehydro
84	129.5	6.7	348	1	S20911	alcohol dehydrogen
85	129	6.7	334	2	E70519	probable oxidoredu
86	128.5	6.7	348	1	DEBYA	alcohol dehydrogen
87	128	6.7	311	2	T51553	quinone oxidoreduc
88	128	6.7	340	2	H95349	alcohol dehydrogen
89	128	6.7	1582	2	E70876	probable polyketid
90	128	6.7	2118	2	S72705	myceroate synth

A;Accession: S59254  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-376 <MAW>  
A;Cross-references: EMBL:Z46843; NID:g861113; PIDN:CAA86891.1; PID:g854503  
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1994  
R;Mallet, L.; Busseteau, F.; Jacquet, M.  
submitted to the Protein Sequence Database, April 1996  
A;Reference number: S63069  
A;Accession: S63079  
A;Molecule type: DNA  
A;Residues: 1-376 <MAW>  
A;Cross-references: EMBL:Z71410; NID:g1302075; PID:e239806; PID:g1302076; MIPS:YNL134C  
A;Experimental source: strain S288C  
C;Genetics:  
A;Cross-references: SGD:S0005078  
A;Map position: 14L

### RESULT 3

62 Q1441 77

db 62 QIVKLGPAVDPKDFSIGDYIYGFIHGSSVRFPSNGAFAEYSAISTVWAYKSPNELKFLGE 121

A; Descrip

**A;Description:** The sequence of *S. cerevisiae* co

```
QY 125 DEISEGPKVNFESASLPVSLTTAGVSLCHHLGSKMEWHPSTPQHTPLLIWGGATAVGQ 184
D 122 DVLPAQVRSLEGAATIPVSLTTAGLVTLNGLNKLWKEPSTPQNGPILLWGGATAVGQ 181
QY 185 QLOVAKHINAYTKIVTVASKKHEKLKSYGADVDVFDYHDAGVIEQIKSKYPNLOHVIDA 244
D 182 SLIQLANLNGFTKIIIVASRKHEKLKKEYGADQLFDYHDIDVVEQIKHKYNNISYLVD 241
QY 245 VGSDDSIPEAYKVTVADSLPATLLEVPMTIESIEEIRKKNVKNKIDITLLYRASQOELLG 304
D 242 VANQNTLQVYKCAADKQDATVLTNLTENVKKNRQNVITDRTLVSIGHEVFPFG 301
QY 305 ATRPPASPEYHEATVKVFPKFNPHLNGDIHNMKNVFSNGLDVPALTEGIEKGNKNV 364
D 302 GITPPADPEARRAATEFVKFNPKISDQIHHPARVVKNGLYDVPRILEDIKIGKNSGE 361
QY 365 KYVARL 370
D 362 KLVAVL 367

RESULT 4
AE1151
oxidoreductase homolog lmo0613 [imported] - Listeria monocytogenes (strain EGD-e)
C;Species: Listeria monocytogenes
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C;Accession: AE1151
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker,
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Faihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Krefte, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend,
A.; Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AE1151
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-313 <GLA>
A;Cross-references: GB:NC_003210; PIDN:CAC98691.1; PID:G16410002; GSPDB:GN00177
A;Experimental source: strain EGD-e
C;Genetics:
C;Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology

Query Match 13.8%; Score 264; DB 2; Length 313;
Best Local Similarity 29.0%; Pred. No. 2.7e-12;
Matches 102; Conservative 56; Mismatches 120; Indels 74; Gaps 18;

QY 8 KAVIIE---GDKAVVKTDSVPPELKEGTALVKVAVAGNPTDWK---HIAYKIGPB-GS 59
D 2 KAVVIENGGKEBEKEVAMPKAGKQNVKEATSNIPDWKLRGKYLKQMDWEFPI 61
QY 60 ILGCDIAGTVVKGPNASTDLKVGDTGFGFVHGASQTDPKNGAFAPAEYARVYPLPYKSNL 119
D 62 ILGWDVAGVISEVGEV-TDMKVGDEVF-----ARPETTRFGTYAETAV-----DDHL 109
QY 120 THSTADEISEGPKVNFESASLPVSLTTAGVSLCHHLGSKMEWHPSTPQHTPLLIWGG 179
D 110 LAPLEPGIS-----FEEAASIPLAGLTAWQALFDH--AKL-----QKGEKVLIHAGA 154
QY 180 TAVGQQLIQVAKHINAYTKIVTVASKKHEKLKSYGADVDVFDYHDAGVIEQIKSKYPN 239
D 155 GGVGTGLAIQLAKYAGA--EVITTSASAKNHELKLSLGADQVLDYKEV---NFKDVLSDID 208
QY 240 HVIDAVGSEDSIPAEYKVTVADSLPATLLEVPMTIESIEEIRKKNVKNKIDITLLYRAS 299
D 209 VFEDTMGGQIE-TDSYDVLKEG-TGRLVSIIVGISNE---DRAKEKNVTAN-GIWLQPN 262
QY 300 EI-----LLGATRFP-----ASPEYHEATVKVFKFN 326
D 263 QLKELGKLLANKTIKPIVGAT-PFPSEKGVFDAHALSETHAVGKIVISFN 312

RESULT 6
T10824
auxin-induced protein (clone MII-3) - mung bean
C;Species: Vigna radiata (mung bean)
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 20-Sep-1999
C;Accession: T10824
R;Chen, J.; Wu, D.; Witham, F.H.; Heuser, C.W.; Arteca, R.N.
submitted to the EMBL Data Library, February 1995
A;Description: Molecular cloning and characterization of auxin-regulated genes from mung
A;Reference number: Z17176
A;Accession: T10824
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-317 <CHE>
A;Cross-references: EMBL:U20808; NID:g1184120; PID:g1184121
A;Experimental source: strain Rwlcz cv. Berken; tissue_type hypocotyl; clone MII-3
C;Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology
```

```
QY 125 DEISEGPKVNFESASLPVSLTTAGVSLCHHLGSKMEWHPSTPQHTPLLIWGGATAVGQ 184
D 122 DVLPAQVRSLEGAATIPVSLTTAGLVTLNGLNKLWKEPSTPQNGPILLWGGATAVGQ 181
QY 185 QLOVAKHINAYTKIVTVASKKHEKLKSYGADVDVFDYHDAGVIEQIKSKYPNLOHVIDA 244
D 182 SLIQLANLNGFTKIIIVASRKHEKLKKEYGADQLFDYHDIDVVEQIKHKYNNISYLVD 241
QY 245 VGSDDSIPEAYKVTVADSLPATLLEVPMTIESIEEIRKKNVKNKIDITLLYRASQOELLG 304
D 242 VANQNTLQVYKCAADKQDATVLTNLTENVKKNRQNVITDRTLVSIGHEVFPFG 301
QY 305 ATRPPASPEYHEATVKVFPKFNPHLNGDIHNMKNVFSNGLDVPALTEGIEKGNKNV 364
D 302 GITPPADPEARRAATEFVKFNPKISDQIHHPARVVKNGLYDVPRILEDIKIGKNSGE 361
QY 365 KYVARL 370
D 362 KLVAVL 367
```



C;Genetics:  
A;Gene: BH0363  
C;Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology

Query Match 10.8%; Score 208; DB 2; Length 324;  
Best Local Similarity 26.4%; Pred. No. 4.6e-08;  
Matches 90; Conservative 53; Mismatches 132; Indels 66; Gaps 15;

Qy 8 KAVIIE---GDKAVKTDVSVPELKEGTALVKVEAVAGNPTDVKHIAKYKPGESIL--- 61  
Db 2 KAIWVTAFGGPEHMMWEDVSTPVYKENEVLKVKVTSNFAD---IKARYGRKGTLFFI 58

Qy 62 -GCDIAGTVVKGPNASTDLKVGDTGFGFVHGASQTDPKNGAFAPYARVYPPIFYKSNLT 120  
Db 59 PGLDAAGYVKGREVS-DIQVGORVIAF-----PKSGSVAEY-----VVAABSLV 103

Qy 121 HSTADEISEGPKVNFESAASLPVSLTTAGVSLCHLGSKMEWHPSTPQHTHPILLINGAT 180  
Db 104 FPIPDEI-----NFRYAAASPI-----VSFLSH-----RLLYNVAQMERGESVLVHAAG 148

Qy 181 AVGQOLIQAQKHINAYTKIVTVASKKHEKLLKSYGADDVDFYHDAGVIEQIKSKYPNLQH 240  
Db 149 GVGTTALQMAKLLGAGTIGTGSADKITAAGSGADEVICYBEEDFTKSV-----QE 201

Qy 241 VIDAVGSEDSIPKAYKVTADSLPATLLEVPVMTIESIPBEIRKDNVKIDI-----TLYRA 296  
Db 202 MTNGVGVD-----IILDSVSGSVTEKSLQCLARYGRLVHFGNSSGAIGTITKIDLHA 253

Qy 297 SGQEIL---LGATFPASPPEVHEATKVFKEINPHLANGDI 334  
Db 254 SCRSLVGLSGTTR-KCKPHLLKETARHVL---PYLASGKL 290

RESULT 10  
C70418  
Probable alcohol dehydrogenase (EC 1.1.1.-) - Aquifex aeolicus  
C;Species: Aquifex aeolicus  
C;Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
C;Accession: C70418  
R;Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; O'V.  
Nature 392, 353-358, 1998  
A;Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.  
A;Reference number: A70300; UID:98196666; PMID:9537320  
A;Accession: C70418  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-343 <AOF>  
A;Cross-references: GB:AE000736; NID:g2983763; PIDN:AAC07327.1; PID:g2983768; GB:AE00065  
A;Experimental source: strain VF5  
C;Genetics:  
A;Gene: adh1  
C;Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology  
C;Keywords: NAD; oxidoreductase  
F;26-332/Domain: long-chain alcohol dehydrogenase homology <LADH>

Query Match 10.1%; Score 193; DB 1; Length 343;  
Best Local Similarity 28.7%; Pred. No. 6.7e-07;  
Matches 77; Conservative 43; Mismatches 94; Indels 54; Gaps 14;

Qy 8 KAVIIEG-----DKAVKTDVSVPELKEGTALVKVEAVAGNPTD-WKHI-AVKIGPE-GSI 60  
Db 2 KAVILTGFGGIEENLKYVEDFPKPEKDEVLIRKVALNHLIDVWRMGALPVKPELPHI 61

Qy 61 LGCDIAGTVVKGPNASTDLKVG-----DTGFGFVHGASQTD-----PKNG 101  
Db 62 LGSVDSVGWEKVG-SLVKVNKEGSEVIAIAPGLSGVCWCQSGRDNHCKDYDILGLKNKG 120

Qy 102 AFAYARVYPPFLFYKSNLTHTADEISEGPKVN--PESASLPVSLTTAGVSLCHLGSK 159  
Db 121 GIABYAV-----PARNVTKPK-KNLSFEEAASYPITFLT-----VWNLVDK 162

Qy 160 MEWHPSTPQHTHPILLINGGATAVGQQLIQVAKHINAYTKIVTVASKKHEKLLKSYGADDV 219

Db 163 AQIKPYS-----RVFIWAGSSGVGAGIQLAKAFNAFV-ITTAGNEEKAKCKELGADLV 216  
Qy 220 FDVHDAGVTEQIKSKY-PNLQHVIVDAVG 246  
Db 217 FHHYKDDVKKRVKREVPKGVGVVDVVDHIG 244

RESULT 11  
G83766  
quinone oxidoreductase BH0935 [imported] - Bacillus halodurans (strain C-125)  
C;Species: Bacillus halodurans  
C;Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001  
C;Accession: G83766  
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, N.; Fujii, F.; Hizen  
Nucleic Acids Res. 28, 4317-4331, 2000  
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and  
A;Reference number: A83650; UID:20512582; PMID:11058132  
A;Accession: G83766  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-322 <STO>  
A;Cross-references: GB:AP001510; GB:BA000004; NID:g10173440; PIDN:BA04654.1; GSPDB:GN001  
A;Experimental source: strain C-125  
C;Genetics:  
A;Gene: BH0935  
C;Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology

Query Match 9.8%; Score 187.5; DB 2; Length 322;  
Best Local Similarity 27.0%; Pred. No. 1.6e-06;  
Matches 79; Conservative 44; Mismatches 83; Indels 87; Gaps 14;

Qy 8 KAVIIE-GDKAVKVT-DVSVPELKEGTALVKVEAVAGNPTD-----WKHIAY 52  
Db 2 KAVVYEQGDPVLKVDVPKPTIGPTDVLINVKASGTPVDVTPRKGRQVQFPHPH 61

Qy 53 KIGPESILGCDIAGTVVKGPNASTDLKVGDTGFGFVHGASQTDPKNGAFAPYARVYPP 112  
Db 62 -----FDVAGEIVEIGSDIE-NMQVGDRVWA-----TNIK-GASAEVALIPEH 102

Qy 113 LPYKSNLTHTADEISEGPKVPESASLPVSLTTAGVSLCHLGSKMEWHPSTPQHTHP 172  
Db 103 LLFP--LPKSV-----SYEEGALAMTPTAHL--L-----FDRGRLOKGET 141

Qy 173 LLIWGATAVGQQLIQVAKHINAYTKIVTVASKKHEKLLKSYGADDVDFVHDAGVIEQIK 232  
Db 142 VLYGSGVAGHNAIQLAKRAGA-TVITTAGREKEGIAKQAGADQVIFKE----- 192

Qy 233 SKYPNLQHVIVDAVGSEDSIPKAYKVTADSLPATLLEVPVMTIE-SIPSEIRKD 284  
Db 193 -----ESVVDVAGK-----ATNGGVPLILDLSLSENMAQD 223

RESULT 12  
AH3214  
zinc-binding oxidoreductase Atu5447 [imported] - Agrobacterium tumefaciens (strain C58; I  
C;Species: Agrobacterium tumefaciens  
C;Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 18-Nov-2002  
C;Accession: AH3214  
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.;  
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McCellig  
; Karp, P.; Romero, P.; Zhang, S.  
Science 294, 2317-2323, 2001  
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, M.  
ster, E.W.  
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
A;Reference number: AB2577; UID:21608550; PMID:11743193  
A;Accession: AH3214  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-334 <KUR>  
A;Cross-references: GB:AE008687; PIDN:AAL46134.1; PID:gl7743901; GSPDB:GN00188  
A;Experimental source: strain C58 (Dupont)

```

Db      175 NTGMFLVOLGKKFGA--KVIAVRK---SWLRGYGADFVVDYNE--VVEKVKKEITNGKMA 227
Qy      237 NLQHVIDAVGSE--DSIPYAKYKTADSLP-ATLLEVVVPMTIESIPBEIRKDNVKIDITLL 293
Db      228 DV--VINSLGEQWDXSFSVLGRKLVITFGTLL-----CGNVKVDLSQL 270
Qy      294 YRASGOEILLGATR 307
Db      271 Y--SKHISILGVNR 282

RESULT 14
Ti2320
alcohol dehydrogenase (EC 1.1.1.1) II - yeast (Candida albicans)
C:Species: Candida albicans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 08-Dec-2000
C:Accession: Ti2320
R:Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, November 1998
A:Reference number: Z18831
A:Accession: Ti2320
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-348 <BAR>
A:Cross-references: EMBL:AL033501; PIDN:CAA21988.1
C:Genetics:
A:Note: adh2
C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology
C:Keywords: alcohol metabolism; NAD; oxidoreductase; zinc
F:44,67,154/Binding site: zinc, catalytic (Cys, His, Cys) #status predicted

Query Match          9.1%; Score 175.5; DB 2; Length 348;
Best Local Similarity 24.6%; Pred. No. 1.4e-05;
Matches 98; Conservative 49; Mismatches 138; Indels 113; Gaps 19;

Qy      1 MSVPTTQKAVIE--GDKAVVKTDSVPELKEGTALVKVEAVAGNPTD---WK---HIAV 52
Db      1 MSVPTTQKAVIFETNGGKLEYK-DIPVPKPKANELLINVKYSGVCHTDLHAWKGDWPLAT 59
Qy      53 KI-----GPEGSILGCDIAGTVKVLGPNASTDLKVG-D-TGGEFVHGA-----SQTDPK 99
Db      60 KLPVLGVGHEG-----AGVVALGENVK-GWKVGDYAGVKVLMGSLNCLNCEYQCQSGAEPN 111
Qy      100 -----NGAFAYEARVYPPLEFYKSNLTHTSTADEISEGVPKVFESAAASLPVSLITTA 148
Db      112 CAEADLSGYTHDGSFQOYA-----TADAV-----QAARIPAGTDLA 147
Qy      149 GVS--LCHHLGSKMWHPTPTQHTHPLLWGGAATVGGQQLIQVAKHINAYTKIVTVASKK 206
Db      148 NVAPILCAGVTYVKALKTAELGAGQWVAISGAAGGLGSLAVQYAKAM-GYRVLAIDGGED 206
Qy      207 HEKLLLSYGADVDYD-HDAGVTEQIKSKYPNLOH-VIDAVGSEDSIPEAYKVTADSLPA 264
Db      207 KGEFVKSLSGAEITFDITKEDVVVEAVKATNGCPGHGVINVSVERAIGOS-----256
Qy      265 TLEVVVPMTIESIPBEIRKDNVKIDITLLYRASGOEILLGATRPFA----SPEYHEATVK 320
Db      257 -----TEYVTLGKVKVLVG----LPAGAKISTVPFVDAVVK 287
Qy      321 FVKFIPNPLNNGDIHNMNIKVFNSGLDDVPALTEGIKE 358
Db      288 TIQIKGSYGVGNKDTAEAVDFFTRGLIKPIKIVGLSE 325

RESULT 15
AG3182
zinc-binding dehydrogenase Atu5188 [imported] - Agrobacterium tumefaciens (strain
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C:Accession: AG3182
R:Wood, D.W.; Secubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.
Gillet, P.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.;
Karp, P.; Romero, P.; Zhang, S.

```





; Karp, P.; Romero, P.; Zhang, S.  
Science 294, 2317-2323, 2001  
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,  
ster, E.W.

A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A:Reference number: AB2577; MUID:21608550; PMID:11743193

A:Accession: AD2699

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-338 <KUR>

A:Cross-references: GB:AE008688; PIDN:AAI42010.1; PID:gl7739384; GSPDB:GN00186

A:Experimental source: strain C58 (Dupont)

C:Genetics:

A:Gene: Atu0996

A:Map position: circular chromosome

Query Match 8.9%; Score 171.5; DB 2; Length 338;  
Best Local Similarity 28.6%; Pred. No. 2.7e-05;  
Matches 64; Conservative 36; Mismatches 89; Indels 35; Gaps 9;

Qy 17 AVVKTDSVPELKEGTALVKVEAVAGNPTDKHIAIKIGPEGS---ILGCDIAGTVVVKLG 73

Db 17 ALIDIDLAQPAKGDHILVEKAVSNVPDVK-VRNQSPENGATRVLGFDSAGVVKAVG 75

Qy 74 PNASTDILKVGDTGF--GFVHGASQTDPKNGAFAYARVYPLFYKSNLTHSTADEISEGP 131

Db 76 DRVSL-FKPGDEVYAGVIN-----RPGSNSEF-----HLVDERIVGAK 113

Qy 132 VK--NFSASALPVSLLTAGVSLCHLGSKNMHPSTPQHTHPLLWGGATAVGQQLIQV 189

Db 114 PKSLNFEEAALPLTAITAYETLFDRLRVK----EPVPGAANAVLVITGGAGGVGSAIQ 169

Qy 190 AKHINAYTKIVTVASKKHEKLLKSYGADDDYDHDAGVIEQIKS 233

Db 170 LRALDLDLTVIATASRPETMEWVKELGAHHVVD-HGKPIAPQVEA 212

RESULT 19

D97481

Probable oxidoreductase (PA3567) [imported] - Agrobacterium tumefaciens (strain C58, Cen

C:Species: Agrobacterium tumefaciens

C:Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 18-Nov-2002

C:Accession: D97481

A:Liiv, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;

R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,

Science 294, 2323-2328, 2001

A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum

A:Reference number: A97359; MUID:21608551; PMID:11743194

A:Accession: D97481

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-338 <KUR>

A:Cross-references: GB:AE007869; PIDN:AAK86805.1; PID:gl5156011; GSPDB:GN00169

C:Genetics:

A:Gene: AGR C 1830

A:Map position: circular chromosome

Query Match 8.9%; Score 171.5; DB 2; Length 338;  
Best Local Similarity 28.6%; Pred. No. 2.7e-05;  
Matches 64; Conservative 36; Mismatches 89; Indels 35; Gaps 9;

Qy 17 AVVKTDSVPELKEGTALVKVEAVAGNPTDKHIAIKIGPEGS---ILGCDIAGTVVVKLG 73

Db 17 ALIDIDLAQPAKGDHILVEKAVSNVPDVK-VRNQSPENGATRVLGFDSAGVVKAVG 75

Qy 74 PNASTDILKVGDTGF--GFVHGASQTDPKNGAFAYARVYPLFYKSNLTHSTADEISEGP 131

Db 76 DRVSL-FKPGDEVYAGVIN-----RPGSNSEF-----HLVDERIVGAK 113

Qy 132 VK--NFSASALPVSLLTAGVSLCHLGSKNMHPSTPQHTHPLLWGGATAVGQQLIQV 189

Db 114 PKSLNFEEAALPLTAITAYETLFDRLRVK----EPVPGAANAVLVITGGAGGVGSAIQ 169

Qy 190 AKHINAYTKIVTVASKKHEKLLKSYGADDDYDHDAGVIEQIKS 233

Db 170 LRALDLDLTVIATASRPETMEWVKELGAHHVVD-HGKPIAPQVEA 212

RESULT 20

AH3289

NADPH2:quinone reductase (EC 1.6.5.5) [imported] - Brucella melitensis (strain 16M)

C:Species: Brucella melitensis

C:Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 03-Jun-2002

C:Accession: AH3289

R:DelVecchio, V.G.; Kaprat, V.; Redkar, R.J.; Patra, G.; Mujet, C.; Los, T.; Ivanova, T.

; Mazur, M.; Goltsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letessier

Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002

A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis

A:Reference number: AD3252; PMID:11756688

A:Accession: AH3289

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-336 <KUR>

A:Cross-references: GB:AE008917; PIDN:AAL51483.1; PID:gl7982195; GSPDB:GN00190

A:Experimental source: strain 16M

C:Genetics:

A:Gene: BMEI0302

A:Map position: I

C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology

C:Keywords: oxidoreductase

Query Match 8.9%; Score 170.5; DB 2; Length 336;  
Best Local Similarity 27.8%; Pred. No. 3.2e-05;  
Matches 68; Conservative 33; Mismatches 101; Indels 43; Gaps 11;

Qy 24 SVPELKEGTALVKVEAVAGNPTDKHIAIKIG-----PEGS--ILGCDIAGTVVVKLG 77

Db 31 AVPEKPEGEILVRRAAGVNRPD---VLQROQGYAPPGASDIPGLEIAGDIVALGHGVK 87

Qy 78 TDLKVGDTGTFGVHGSQTDPKNGAFAYARVYPLFYKSNLTHSTADEISEGPVKNFES 137

Db 88 R-FRKGDQVALLAG-----GGYAEYAVH-----ESNALPLSG--FGYIE 126

Qy 138 AASLPVSLTTAGVSLCHLGSKNMHPSTPQHTHPLLWGGATAVGQQLIQVAKHINAYT 197

Db 127 AAAPETFTFVHNVFERGLK-----EGEVLVHVGSSGIGTTAIQAKAFGA-T 176

Qy 198 KIVTVASKKHEKLLKSYGADDDYDHDAGVIEQIKSKY--PNLQHVIDAVGSDSDSPEAY 255

Db 177 VITTAGSKKCDACVKLGADRAINHEEDFVAVVKEMTGKGVVDVILDMVGG--DYVERN 235

Qy 256 KVTAD 260

Db 236 KAAAE 240

RESULT 21

T10203

Hypothetical protein F25G13.100 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cross)

C:Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 15-Oct-1999

C:Accession: T10203

R:Bevan, M.; Pohl, T.; Weizensegger, T.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.; Lemcke,

submitted to the Protein Sequence Database, June 1999

A:Reference number: Z16991

A:Accession: T10203

A:Molecule type: DNA

A:Residues: 1-329 <BEV>

A:Cross-references: EMBL:AL079349; GSPDB:GN00062; ATSP:F25G13.100

A:Experimental source: cultivar Columbia; BAC clone F25G13

C:Genetics:

A:Gene: ATSP.F25G13.100

A:Map position: 4

A:Introns: 22/3; 70/1; 100/3

C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology



ster, E.W.

A>Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A;Reference number: AB2577; MUID:21608550; PMID:11743193

A;Accession: A12677

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-327 <KUR>

A;Cross-references: GB:AB008688; PIDN:AAL41839.1; PID:g17739198; GSPDB:GN00186

A;Experimental source: strain C58 (Dupont)

C;Genetics:

A;Gene: qor

A;Map position: circular chromosome

C;Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology

Query Match 8.4%; Score 162; DB 2; Length 327;  
Best Local Similarity 24.0%; Pred. No. 0.00013;  
Matches 86; Conservative 57; Mismatches 131; Indels 84; Gaps 17;

QY 14 GDKAVVKTDSVPPELKEGTALVKVEAVAGNPTDWMKHIAYKIG-----PEGS--ILGCDIAG 67  
DB 11 GPEVMQSSKAPLPKFPARGELILKVEAGVNPDP---VAQRQGIYPPPKGASPLGLEIAG 67  
QY 68 TVVKLGNASTDLKVGDTGFGFVHGASQTDPKNGAFAYARVYPPPLFYKSNLTHSTADEI 127  
DB 68 EEWALGEGVD-EFKLGDVKVCALANG-----GGYAEYCAV-----PAGQA 105  
QY 128 SEGPKNFES--AASLPVSLTTAGVSLCHLGSKMEWHPTPQHTHPLLIWGATAVGQQ 185  
DB 106 LPFP-KGYDVAKAAALPETFTFWANLFWAG-----LTEGET--VLHGTSGIGTT 155  
QY 186 LIQVAKHINAYTKIVTAVSKHEKLKSYGADDFVYHDAGVIEQIKSKY--PNLQHVLD 243  
DB 156 AIQLAKAFGEVAYATAGSAEACVK-LGTYKRAINREDPFAIVKSEFGCKGVVDVLD 214  
QY 244 AVGSEDSIPEAYKVTADSLPATLLEVVPMTIESIPEIRKDNVKIDITLYRASGQEI-- 301  
DB 215 MIGA-----AY-----FEKLAALAKDGCLSLIARLGGATAEKVDL 250  
QY 302 -----LLGATFPASPPEYHEA-TVKPKVFINPLNNGDIHHMKIVFNSGLDDV 349  
DB 251 RPIMVKRLTVGTSMRPTADEKRAIRDELVEQWPLIESGKAPVINRVFT--LEEV 306

#### RESULT 25

A70871

Probable quinone oxidoreductase - Mycobacterium tuberculosis (strain H37Rv)

C;Species: Mycobacterium tuberculosis

C;Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000

C;Accession: A70871

R;Cole, S.T.; Broesch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.

; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A;Reference number: A70500; MUID:98295987; PMID:9634230

A;Accession: A70871

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-328 <COL>

A;Cross-references: GB:AL021184; GB:AL123456; NID:g3261498; PIDN:CAA15984.1; PID:g279138

A;Experimental source: strain H37Rv

C;Genetics:

A;Gene: qor

C;Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology

F;25-318/Domain: long-chain alcohol dehydrogenase homology <LADH>

Query Match 8.4%; Score 161.5; DB 2; Length 328;  
Best Local Similarity 29.5%; Pred. No. 0.00015;  
Matches 72; Conservative 36; Mismatches 95; Indels 41; Gaps 13;

QY 14 GDKAVVK-TDVSPELKEGTALVKVEAVAGNPTDWMKHIAYKIGPE-GSILGCDIAGTVVK 71

DB 10 GGPGLRHRVDQPOPOPGHCELLIKABEIGAIVNFIDTYFRSQYPRELPFVIGSEVCQGTVEA 69  
QY 72 LGPN---ASTDLKVGDTGFGFVHGASQTDPKNGAFAYARVYPPPLFYKSNLTHSTADEIS 128  
DB 70 VGPVTAADTAISVGDV---VVSAS---ANGAYAEFCTA-----PASLTAKVPDDVT 115  
QY 129 EGPVKNFESAASLPVSLTTAGVSLCHLGSKMEWHPTPQHTHPLLIWGATAVGQOLIQ 188  
DB 116 S-----EVAASALLKGLTA-----HYLLKSV--YPVKEGDT--VLVHAGAGGVGLILTO 160  
QY 189 VAKHINAYTKIVTAVSKHEKLKSYGADDFVYHD-----AGVIEQIKSKYPNLQHVLD 243  
DB 161 WATHLGRV-ITTVSTAERAKLSKADAGADVLDYPEDAWQFAGRVELTGG-TGVOAVVD 218  
QY 244 AVGS 247  
DB 219 GVGA 222

#### RESULT 26

PN0448

zeta-crystallin / quinone reductase (NADPH) (EC 1.6.-.-) - human

C;Species: Homo sapiens (man)

C;Date: 31-Dec-1993 #sequence\_revision 07-Jul-1995 #text\_change 11-Jun-1999

C;Accession: PN0448; A54672

R;Gonzalez, P.; Rao, P.V.; Zigler Jr., J.S.

Biochem. Biophys. Res. Commun. 191, 902-907, 1993

A;Title: Molecular cloning and sequencing of zeta-crystallin/quinone reductase cDNA from

A;Reference number: PN0448; MUID:93221534; PMID:8466529

A;Accession: PN0448

A;Molecule type: mRNA

A;Residues: 1-329 <CON>

A;Cross-references: GB:S58039; NID:g299369; PIDN:AAB26039.1; PID:g299370

A;Experimental source: liver

A;Note: translation of initiator Met is not shown; the authors translated the codon ATC to

R;Gonzalez, P.; Rao, P.V.; Zigler Jr., J.S.

Genomics 21, 317-324, 1994

A;Title: Organization of the human zeta-crystallin/quinone reductase gene (CRY2).

A;Reference number: A54672; MUID:94375054; PMID:8088825

A;Accession: A54672

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-37 <GO2>

A;Cross-references: GB:L31521

C;Comment: This protein is present at low (enzymatic) levels in this species, in contrast

#### C;Genetics:

A;Gene: GDB:CRY2

A;Cross-references: GDB:139194; OMIM:123691

A;Map position: lp31-lp22

C;Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology

C;Keywords: eye lens; NAD; oxidoreductase

F;33-318/Domain: long-chain alcohol dehydrogenase homology <LADH>

F;152-181/Region: beta-alpha-beta NADP nucleotide-binding fold

Query Match 8.4%; Score 161; DB 1; Length 329;

Best Local Similarity 24.8%; Pred. No. 0.00016;

Matches 73; Conservative 50; Mismatches 95; Indels 76; Gaps 14;

QY 19 VKTDSVPPELKEGTALVKVEAVAGNPTDWMKHIAYKIGPEGSIL-----GCDIAGTVVKLGP 74

DB 24 LRSADIAPIPKDHQVLILKIVHACGVNPE-TYIRSGYTSRKPLLPYTPGSDVAGVIEAVGD 82

QY 75 NASTDLKVGDTGFGFVHGASQTDPKNGAFAYARVYPPPLFYKSNLTHSTADEISEGPVK- 133

DB 83 NASA-FKGDVPE-----TSSTISGGYAEYALAADHTVYKL-----PEKL 121

QY 134 NFESAASLPVSLTTAGVSLCHLGSKMEWHPTPQHTHPLLIWGATAVGQOLIQVAKHI 193

DB 122 DFKQGAIGIPYFTAYRALIHSAVKA-----GESVLVHGASGGVGLAACQIAR-- 170

QY 194 NAY-TKIV-TVASKKHEKLKSYGADDFVYHDAGVIEQIKSKYPNLQHVLDVAVGSEDSI 251

Db 171 -AYGLKILGTAGTEGQKVLQNGAHEVFNHREVNYYIDKIK-KY-----VGEKG-- 217

QY 252 PEAYKVTADSLPATLLEVVPMTIESPEIRKDNVKIDITLLYRASGOEILLGA 305

Db 218 -----IDIIEMLANVLSKDLULSH-GGRVIVGS 248

RESULT 27

AC1858

hypothetical protein all0412 [imported] - Nostoc sp. (strain PCC 7120)

C;Species: Nostoc sp. PCC 7120

A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C;Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Dec-2002

C;Accession: AC1858

R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.

DNA Res. 8, 205-213, 2001

A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena

A;Reference number: AB1807; MUID:2159285; PMID:11759840

A;Accession: AC1858

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-318 <CUR>

A;Cross-references: GB:BA000019; PIDN:BA072370.1; PID:g17129757; GSPDB:GN00179

A;Experimental source: strain PCC 7120

C;Genetics:

C;Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology

Query Match 8.3%; Score 159.5; DB 2; Length 318;

Best Local Similarity 30.0%; Pred. No. 0.0002;

Matches 63; Conservative 25; Mismatches 87; Indels 35; Gaps 7;

QY 20 KTDVSVPELKEGTALVKVEAVAGNPTDKHIAKIGPE---GSTLGCIDIACTVVKLGPNA 76

Db 17 EVENKPTPANNEVLVKVYATSNPACDGMRGFFGRVQLPALGLDVSQVVEAVGENV 76

QY 77 STDCLKVGTGFGFVHGASQTDPKNGAPAEYARVYVPLFYK--SNLTHSTADEISEGVPKN 134

Db 77 -RDFQVGDE---VYVAIPHELGGANAEYHAPSSMAKKPNNSHL----- 119

QY 135 FESAASLPVSLTAGVSLCHLGSKMEWHPSTPQHTPLLIWGGATAVGQQLIQVAKHI 194

Db 120 --BAATVPVAGGTAAWALITR-----ANLQVGETVLHGGAGGVGTFAIQAKAG 168

QY 195 AYTKIVTVASKKHEKLLKSYGADDPDYHD 224

Db 169 AY--VYTCGGYDIDFVKSGADRAIDYRN 196

RESULT 28

CYGPZ

zeta-crystallin / quinone reductase (NADPH) (EC 1.6.-.-) - guinea pig

N;Alternate names: quinone reductase (NADP)

C;Species: Cavia porcellus (guinea pig)

C;Date: 31-Mar-1990 #sequence\_revision 31-Mar-1990 #text\_change 11-Jun-1999

C;Accession: J50230

R;Rodokanaki, A.; Holmes, R.K.; Borraas, T.

Gene 78, 215-224, 1989

A;Title: Zeta-crystallin, a novel protein from the guinea pig lens is related to alcohol

A;Reference number: J50230; MUID:89378748; PMID:2777081

A;Accession: J50230

A;Molecule type: mRNA

A;Residues: 1-329 <RD>

A;Cross-references: GB:M26936; NID:g191252; PIDN:AAA37035.1; PID:g305333

A;Experimental source: strain 13/N

A;Note: the sequences of seven fragments (79-98, 95-106, 106-114, 209-215, 264-275, 291-

C;Comment: This protein is distantly related to mammalian and yeast alcohol dehydrogenases

C;Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology

C;Keywords: eye lens; NADP; oxidoreductase

F;152-181/Region: beta-alpha-beta NADP nucleotide-binding fold

Query Match 8.3%; Score 159; DB 1; Length 329;

Best Local Similarity 24.2%; Pred. No. 0.00023;

Matches 71; Conservative 50; Mismatches 96; Indels 76; Gaps 14;

QY 19 VKTDSVPPELKEGTALVKVEAVAGNPTDKHIAKIGPEGSIL-----GCDIAGTVVKLGP 74

Db 24 VQSDVAVPPIKQHQVLKIVHACINPVE-TYIRSGTYTRIPLLPYTPGTVDVAGVBSIGN 82

QY 75 NASTDLKVGDTGFGFVHGASQTDPKNGAPAEYARVYVPLFYKSNLTHSTADEISEGVPK- 133

Db 83 DVSA-FKKGRVLP-----TTSTISGGYAEVALASDHTVYRL-----PEKL 121

QY 134 NFESAASLPVSLTAGVSLCHLGSKMEWHPSTPQHTPLLIWGGATAVGQQLIQVAKHI 193

Db 122 DFRQGAAGIPYPTA---CRAL-----FHSARAKAGESVLVHGASGGVGLAACQIAR-- 170

QY 194 NAY-TKIV-TVASKKHEKLLKSYGADDPDYHDVAGVIEQIKSKYPNLQHVIVDAVGSEDSI 251

Db 171 -AYGLKVLGTAGTEGQKVLQNGAHEVFNHRAHYIDEIKK-----SIGEKG-- 217

QY 252 PEAYKVTADSLPATLLEVVPMTIESPEIRKDNVKIDITLLYRASGOEILLG 304

Db 218 -----VDVIIEMLANVLSNDLKL--SCGGRVIVG 247

RESULT 29

AS4932

zeta-crystallin / quinone reductase (NADPH) (EC 1.6.-.-) - mouse

C;Species: Mus musculus (house mouse)

C;Date: 25-Apr-1995 #sequence\_revision 07-Jul-1995 #text\_change 11-Jun-1999

C;Accession: AS4932; S42273

R;Gonzalez, P.; Hernandez-Calzadilla, C.; Rao, P.V.; Rodriguez, I.R.; Ziegler Jr., J.S.; I

Mol. Biol. Evol. 11, 305-315, 1994

A;Title: Comparative analysis of the zeta-crystallin/quinone reductase gene in guinea pi

A;Reference number: AS4932; MUID:94224126; PMID:8170370

A;Accession: AS4932

A;Molecule type: mRNA

A;Residues: 1-331 <CON>

A;Cross-references: GB:S70056; NID:9546493; PIDN:AAB30620.1; PID:9546494

A;Experimental source: liver

A;Note: Sequence extracted from NCBI backbone (NCBIN:147625, NCBIPI:147626)

A;Note: Translation of initiator Met is not shown

R;Joernvall, H.; Persson, B.; du Bois, G.C.; Laveza, G.C.; Chen, J.H.; Gonzalez, P.; Rao,

FEBS Lett. 322, 240-244, 1993

A;Title: zeta-Crystallin versus other members of the alcohol dehydrogenase super-family.

A;Reference number: S42272; MUID:93252077; PMID:8486156

A;Contents: annotation

C;Comment: This protein is a major soluble protein of the lens in this species and is exi

C;Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology

C;Keywords: eye lens; NADP; oxidoreductase

F;33-320/Domain: long-chain alcohol dehydrogenase homology <LADH>

F;152-181/Region: beta-alpha-beta NADP nucleotide-binding fold

Query Match 8.3%; Score 158.5; DB 1; Length 331;

Best Local Similarity 22.4%; Pred. No. 0.00025;

Matches 72; Conservative 53; Mismatches 119; Indels 77; Gaps 13;

QY 19 VKTDSVPPELKEGTALVKVEAVAGNPTDM--KHIAKIGPEGSIL-GCDIAGTVVKLGP 75

Db 24 LQSDVVVPVQSHQVLKIVHACINPVEVTVIRSGYSRKPALPYTPGSDVAGVIESVGDK 83

QY 76 ASTDLKVGDTGFGFVHGASQTDPKNGAPAEYA-----RVYVPLFYKSNLTHSTADEISEG 130

Db 84 VSA-FKKGRVFCY-----STVSGGYAEFALAADDTIYF-----LPE- 119

QY 131 PVKNFESAASLPVSLTAGVSLCHLGSKMEWHPSTPQHTPLLIWGGATAVGQQLIQVA 190

Db 120 -TLNFRQGAALGIPYPTA---CRAL-----FHSARAKAGESVLVHGASGGVGLATQIA 169

QY 191 KHTNAYTKIVTVASKKHEKLLKSYGADDPDYHDVAGVIEQIKSKYPNLQHVIVDAVGSEDS 250

Db 170 R-AHGLKVLGTAGSEGGKVLQNGAHEVFNHKEANYIDIKM-----SVGDKDK 218



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 10, 2004, 08:41:33 ; Search time 23 Seconds  
(without alignments)  
756.517 Million cell updates/sec

Title: US-10-081-644-2

Perfect score: 1320

Sequence: 1 MSVPTQKAVIEGDKAVK.....ALTEGIKEGKNVYVARL 370

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 90 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Query Length	ID	Description
1	1065.5	55.5	376	1 YNN4 YEAST	P3912 saccharomyc
2	1021.5	53.2	376	1 YL60 YEAST	P54007 saccharomyc
3	1015.5	52.9	368	1 YCZ2 YEAST	P25608 saccharomyc
4	370.5	19.3	297	1 TOXD COCCA	P54006 cochlidiobol
5	254	13.2	309	1 QORL ARATH	Q84001 arabidopsis
6	175.5	9.1	348	1 ADH2 CANAL	Q84038 candida alb
7	161	8.4	329	1 QOR HUMAN	Q82527 homo sapien
8	159	8.3	329	1 QOR CAVPO	P11415 cavia porce
9	158.5	8.3	331	1 QOR MOUSE	P47199 mus musculu
10	155	8.1	340	1 QOR LEIAM	P28625 leishmania
11	154	8.0	365	1 YM27 YEAST	P28625 saccharomyc
12	153.5	8.0	345	1 PL ARATH	Q39172 arabidopsis
13	153.5	8.0	348	1 ADH1 KLUMA	Q07288 kluyveromyc
14	152.5	7.9	348	1 ADH2 PICST	Q33039 pichia stip
15	149.5	7.8	348	1 ADH1 PICST	O00097 pichia stip
16	148	7.7	330	1 QOR BOVIN	O97764 bos taurus
17	148	7.7	330	1 QOR LANGU	Q28452 lama guanac
18	146	7.6	343	1 P2 ARATH	Q39173 arabidopsis
19	145.5	7.6	350	1 ADH1 KLUMA	P20369 kluyveromyc
20	143.5	7.5	350	1 ADH1 CANAL	P43067 candida alb
21	142.5	7.4	348	1 TDH PYRHO	O58389 pyrococcus
22	142	7.4	341	1 TDH ECOL6	Q8fca2 escherichia
23	141.5	7.4	337	1 ADH1 BACST	P12311 bacillus st
24	141.5	7.4	348	1 TDH PYRAB	Q9uyx0 pyrococcus
25	141	7.3	341	1 TDH ECOS7	Q8xej1 escherichia
26	141	7.3	341	1 TDH SHIFL	P59409 shigella fl
27	140.5	7.3	347	1 ADH2 KLUMA	Q9p4c2 kluyveromyc
28	140	7.3	341	1 TDH ECOLI	P07913 escherichia
29	140	7.3	341	1 TDH SHEON	Q8e8j1 shewanella
30	140	7.3	379	1 VATI TORCA	P93333 torpedo cal
31	137.5	7.2	348	1 TDH PYRPU	Q8u259 pyrococcus
32	134.5	7.0	334	1 ADH3 KLUMA	P49384 kluyveromyc
33	132	6.9	334	1 QOR YEAST	P38230 saccharomyc

#### ALIGNMENTS

##### RESULT 1

YNN4_YEAST	YNN4_YEAST	STANDARD;	PRT;	376 AA.
ID	YNN4_YEAST			
AC	P53912;			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	01-OCT-1996 (Rel. 34, Last annotation update)			
DE	Hypothetical 41.2 kDa protein in PP1-TOM22 intergenic region.			
GN	YNN134C OR N1214 OR N1847.			
OS	Saccharomyces cerevisiae (Baker's yeast).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			

P07246 saccharomyc  
Q06004 bacillus su  
Q02912 bombyx mori  
P49383 kluyveromyc  
Q31186 rhizobium m  
P42328 bacillus st  
Q822f4 salmonella  
Q82152 salmonella  
P00331 saccharomyc  
P00330 saccharomyc  
Q30359 nicotiana t  
Q9k162 vibrio chol  
O82515 medicago sa  
P27867 rattus norv  
P42327 bacillus st  
P59410 vibrio para  
Q38707 apium grave  
P38113 saccharomyc  
Q99536 homo sapien  
P54202 emericeella  
P39451 escherichia  
P50381 sulfolobus  
P30360 nicotiana t  
Q64442 mus musculu  
Q82jnz2 yersinia pe  
P31656 medicago sa  
Q31776 bacillus su  
P00332 schizosacch  
Q91233 streptomyces  
Q52998 rhizobium m  
P39462 sulfolobus  
Q45697 caenorhabdi  
P42734 arabidopsis  
P80715 streptococc  
Q00736 homo sapien  
P38105 escherichia  
Q49482 arabidopsis  
Q57517 haemophilus  
P31657 populus del  
Q04894 saccharomyc  
P19212 neutrospora  
Q92rri1 fragaria an  
P35630 entamoeba h  
Q03132 saccharopol  
P39450 pasteurella  
P07846 ovis aries  
Q8d442 vibrio vuln  
P72334 rhodobacter  
P40783 salmonella  
Q98317 rhizobium l  
Q9k714 bacillus ha  
Q02972 arabidopsis  
P25437 escherichia  
Q43137 styloanthus  
P25377 saccharomyc  
Q10094 schizosacch  
P42754 petroselinu

```

OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C; PubMed=8619318;
RX MEDLINE=96109932; Bussereau F., Jacquet M.;
RA "A 43.5 kb segment of yeast chromosome XIV, which contains MFA2,
RT MEP2, CAP/SRV2, NAM9, FKBI/FPB1/RBP1, MOM22 and CPT1, predicts an
RT adenosine deaminase gene and 14 new open reading frames.";
RL Yeast 11:1195-1209(1995).
CC -!- SIMILARITY: BELONGS TO THE YCR102C/YLR460C/YNL134C FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z46843; CRA86891.1; -.
DR EMBL; Z71410; CRA96016.1; -.
DR PIR; S55149; S55149.
DR SGD; S0005078; YNL134C.
DR InterPro; IPR002085; Adh_zn_family.
DR Pfam; PF00107; ADH_zinc_N; 1.
KW Hypothetical protein.
SQ SEQUENCE 376 AA; 41164 MW; AE39BBBCDA46C3B9 CRC64;

Query Match 55.5%; Score 1065.5; DB 1; Length 376;
Best Local Similarity 54.0%; Pred. No. 1.4e-69;
Matches 201; Conservative 59; Mismatches 109; Indels 3; Gaps 2;

QY 2 SVPTQKAVIIEGDKAVVKTDSVPELKEGTALVKVEAVAGNPTDWHKIAVKGIGESI 61
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4 SIPETMKAVIENGKAVVQDIPPELEEGFVLKTVAVAGNPTDWHKIDFKIGPQCALL 63

QY 62 GCDIAGTVVVKLGPNA-STDLVKGTGFGFVHGASQTDPKNGAFAYARVPPFYK--SN 118
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
64 GCDAAQGVKLGPNVDARFAIGDIYGVHIGASVRFPSNGAFAYSAISSETAYKPARE 123

QY 119 LTHSTADISEGPKVKNFESAASLPVSLTGTAGVSLCHLHLSGKMEWHPTPOHTPLIWWG 178
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
124 FRLCGKDKLPSPVKVSLGAVSLVSTAGMLTHTSFGLDMTWKPKAQDQPILEWGG 183

QY 179 ATAVGQQLIQVAKHINAYTKITVTVASKKHEKLSYGADVDVFDHAGVIEQIKSKYPNL 238
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
184 ATAVGQMLIQALKLNGFSKLIIVASRKHEKLLKEYGADELFDYHDADVIEQIKKYNNI 243

QY 239 QHVDVAGSEDSIPEAYKVTDADSLPATLLEVVVPMTIESIPEIRKDNVKIDITLLYRASG 298
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
244 PYLVDCVSNTEIIQQVYKCAADDATVQVLTVLTEKDIKEEDRRQNVSEGTLLYLG 303

QY 299 QEILLAGTRFPASPEYHEATVKVFKNPHLNGDIHNMNKKVFNGLDDVPALTEGKE 358
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
304 NDVPFGFTFLPADPEYKAAIKFKINPKINDGQIRHPIVVRVYKNGCLDVPALDKIKH 363

QY 359 GKNKNVYKVARL 370
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
364 GRNSGEKLVAVL 375

RESULT 2
YL60 YEAST STANDARD; PRT; 376 AA.
AC P54007;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Hypothetical 41.1 kDa protein ON CDC91-PAU4 intergenic region.
GN YLR460C OR L9122.7.
OS Saccharomycetes cerevisiae (Baker's yeast).

```

```

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OX Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
RN NCBI_TaxID=4932;
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=97313267; PubMed=9169871;
RA Johnston M., Hillier L., Riles L., Albermann K., Andre B., Ansorge W.,
RA Benes V., Brueckner M., Delius H., Dubois E., Duesterhoeft A.,
RA Entian K.-D., Floeth M., Goffeau A., Hebling U., Heumann K.,
RA Heuss-Neitzel D., Hilbert H., Hilger F., Kleine K., Koetter P.,
RA Louis E.J., Messenguy F., Mewes H.-W., Miosga T., Moestl D.,
RA Mueller-Auer S., Nentwich U., Obermaier B., Piravandi E., Pohl T.M.,
RA Portetelle D., Purnelle B., Rechmann S., Rieger M., Rinke M., Rose M.,
RA Scharfe M., Scherens B., Scholler P., Schwager C., Schwarz S.,
RA Underwood A.P., Urreazaru L.A., Vandenbol M., Verhasselt P.,
RA Vierendeels F., Voet M., Voickaert G., Voss H., Wambutt R., Wedler E.,
RA Wedler H., Zimmermann F.K., Zollner A., Hani J., Hoheisel J.D.;
RT "The nucleotide sequence of Saccharomycetes cerevisiae chromosome XII.";
RL Nature 387:87-90(1997).
CC -!- SIMILARITY: BELONGS TO THE YCR102C/YLR460C/YNL134C FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U22383; AAB64723.1; -.
DR PIR; S59418; S59418.
DR SGD; S0004452; YLR460C.
DR InterPro; IPR002085; Adh_zn_family.
DR Pfam; PF00107; ADH_zinc_N; 1.
KW Hypothetical protein.
SQ SEQUENCE 376 AA; 41127 MW; 00139949423862F1 CRC64;

Query Match 53.2%; Score 1021.5; DB 1; Length 376;
Best Local Similarity 51.7%; Pred. No. 2.1e-66;
Matches 193; Conservative 67; Mismatches 110; Indels 3; Gaps 2;

QY 1 MSVPTQKAVIIEGDKAVVKTDSVPELKEGTALVKVEAVAGNPTDWHKIAVKGIGESI 60
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3 VAIPETMKAVIIEGDKAVVKTDSVPELKEGTALVKVEAVAGNPTDWHKIAVKGIGESI 62

QY 61 LGCDIAGTVVVKLGPNA-STDLVKGTGFGFVHGASQTDPKNGAFAYARVPPFYK-- 117
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
63 LGCDAAQGVKLGPAVNPDKFDSIGDIYGVHIGASVRFPSNGAFAYSAISTVAYKSPN 122

QY 118 NLTHSTADISEGPKVKNFESAASLPVSLTGTAGVSLCHLHLSGKMEWHPTPOHTPLIWWG 177
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
123 ELKFLGEDVLPAGVRSLEGVATIPVSLTGTAGVSLCHLHLSGKMEWHPTPOHTPLIWWG 182

QY 178 GATAVQQLIQVAKHINAYTKITVTVASKKHEKLSYGADVDVFDHAGVIEQIKSKYPN 237
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
183 GATAVQQLIQVAKHINAYTKITVTVASKKHEKLSYGADVDVFDHAGVIEQIKSKYPN 242

QY 238 LQHVIDVAGSEDSIPEAYKVTDADSLPATLLEVVVPMTIESIPEIRKDNVKIDITLLYRAS 297
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
243 ISYLVDCVANQDTLQQVYKCAADKQDQATIVELKNTLEENVKKNRQNVTDIIRLYSIG 302

QY 298 QSBILLGATRFAPASPEYHEATVKVFKNPHLNGDIHNMNKKVFNGLDDVPALTEGKE 357
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
303 GHEVFPFGNITLPADSEARAAIKFKINPKINDGQIRHPIVVRVYKNGCLDVPALDKIK 362

QY 358 EGKKNVYKVARL 370
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
363 YGNSGEKLVAVL 375

RESULT 3
YCZ2_YEAST

```





RA White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,  
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,  
RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,  
RA Dunn P., Egu P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y.,  
RA Giller J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,  
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,  
RA Kim C.J., Koo H.L., Krenetskaia I., Kurtz D.B., Kwan A., Lam B.,  
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,  
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A.,  
RA Miltscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,  
RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,  
RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,  
RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,  
RA Uterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,  
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W., of the plant Arabidopsis  
RT thaliana.";  
RL Nature 408:816-820(2000).  
CC -!- SIMILARITY: Belongs to the zinc-containing alcohol dehydrogenase  
CC family. Quinone oxidoreductase subfamily.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AC005990; AAC98029.1; -.  
DR PIR; E86371; E86371.  
DR SWISS-2DPAGE; Q9TUC1; ARATH.  
DR InterPro; IPR002085; Adh\_zn family.  
DR InterPro; IPR002364; QOR\_zeta\_crystal.  
DR Pfam; PF00107; ADH\_zinc\_N; 1.  
DR PROSITE; PS01162; QOR\_ZETA\_CRYSTAL; FALSE\_NEG.  
KW Oxidoreductase; Zinc; NAD.  
SQ SEQUENCE 309 AA; 32775 MW; EC328042771BEE6 CRC64;  
  
Query Match 13.2%; Score 254; DB 1; Length 309;  
Best Local Similarity 32.8%; Pred. No. 3.3e-11;  
Matches 84; Conservative 35; Mismatches 97; Indels 40; Gaps 10;  
  
QY 15 DKAVKTDVSVPELKEGTALVKVAVAGNPTDMKHIAKYI----GPGSILGCDIAGTVV 70  
DB 13 DVLKESINIVPEIKEDQVLKVAALNPVDKRRQCKFKATSDPLTPGVDVAGVV 72  
QY 71 KLGPNASTDLLKVGDTGFGFVHGASQTOPKN-GAPAEYARVYVPL--FYKSNLTHSTADEI 127  
DB 73 KVG-SAVKDLKEGVGVANVSEKALEGPKQFGSLAEYTAVEKLLALPKNI----- 123  
QY 128 SEGPKVPESAAISLPSVLTAGVSLCHHLGSKMWHPTSPQTHPLLIWGATVAGQOLI 187  
DB 124 -----DPAQAGLPLAETADEGLV-----RTEFSAG-----KSILVLNAGGVGSLVI 167  
QY 188 QVAKHINAYTKVTVVASKHEKLLKSYGADDDVFDHAGVTEIQISKYPNLQHVIVDAVGS 247  
DB 168 QLAKHVYGAASKAVATASTEKLELVRSLGALDAIDYTKEN-IEDLPDKY----DVVFDAGM 223  
QY 248 EDS-----IPEAYKVTA 259  
DB 224 CDKAVKIKEGGKVA 239  
  
RESULT 6  
ADH2 CANAL STANDARD; PRT; 348 AA.  
AC O94038;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Alcohol dehydrogenase 2 (EC 1.1.1.1).  
GN ADH2 OR CA41C10.04.

OS Candida albicans (Yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.  
OX NCBI\_TaxID=5476;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=1161;  
RA Taylor K., Harris D., Barrell B.G., Rajandream M.A.;  
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
CC -!- CATALYTIC ACTIVITY: An alcohol + NAD(+) = an aldehyde or ketone +  
CC NADH.  
CC -!- COFACTOR: Binds 2 zinc ions per subunit (By similarity).  
CC -!- SUBUNIT: Homotetramer (By similarity).  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).  
CC -!- SIMILARITY: Belongs to the zinc-containing alcohol dehydrogenase  
CC family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AL033501; CAA21988.1; -.  
DR PIR; T18230; T18230.  
DR InterPro; IPR002328; ADH\_zinc.  
DR InterPro; IPR002085; Adh\_zn family.  
DR Pfam; PF00107; ADH\_zinc\_N; 1.  
DR PROSITE; PS00059; ADH\_ZINC; 1.  
KW Oxidoreductase; Zinc; Metal-binding; NAD.  
FT METAL 44 44 ZINC 1 (CATALYTIC) (BY SIMILARITY).  
FT METAL 67 67 ZINC 1 (CATALYTIC) (BY SIMILARITY).  
FT METAL 98 98 ZINC 2 (BY SIMILARITY).  
FT METAL 101 101 ZINC 2 (BY SIMILARITY).  
FT METAL 104 104 ZINC 2 (BY SIMILARITY).  
FT METAL 112 112 ZINC 2 (BY SIMILARITY).  
FT METAL 154 154 ZINC 1 (CATALYTIC) (BY SIMILARITY).  
SQ SEQUENCE 348 AA; 36807 MW; 3D6B6753852AD2DCF CRC64;  
  
Query Match 9.1%; Score 175.5; DB 1; Length 348;  
Best Local Similarity 24.6%; Pred. No. 1.7e-05;  
Matches 98; Conservative 49; Mismatches 138; Indels 113; Gaps 19;  
  
QY 1 MSVPTTQKAVIE--GDKAVVKTDSVPELKEGTALVKVAVAGNPTD--WK---HIAY 52  
DB 1 MSVPTTQKAVIFETNGGKLEYK-DIPVPKPKANELLINVKYSGVCHTDLHAWKGDWPLAT 59  
QY 53 KI----GPGSILGCDIAGTVVKGPNASTDLKVGDTGFGFVHGA-----SQTDPK 99  
DB 60 KLPLVGGHEG-----AGVVALGENVK-GWKVGYAGVKNLNGSCLNCEYCCSGAEPN 111  
QY 100 -----NGAFAYEARVYVPLFYKSNLTHSTADEISEGPKVKNFESAASLPVSLTTA 148  
DB 112 CAEADLSGYTHDGSFQYYA-----TADAV-----QAARIPAGTDLA 147  
QY 149 GVS--LCCHLGSKMWHPTSPQTHPLLIWGATVAGQOLIQVAKHINAYTKVTVVASKK 206  
DB 148 NVAPILCAGTVVYKALKTALEAGQVVAISGAAGGLGSLAVQYAKAM-GYRVLAIIDGGED 206  
QY 207 HEKLLKSYGADDDVFDY-HDAGVTEIQISKYPNLQHVIVDAVSEDSIPEAYKVTVADSLPA 264  
DB 207 KGEFVKSLGAETFDITFTEKQDVVEAVKKAATNGGPGHGVINVSVERAIGS----- 256  
QY 265 TLLEVVPMTIESPEIRKDNVKNIDITLLYRASQOILLGATRFPA-----SPEYHEATVK 320  
DB 257 -----TEVYRTLKGVLVG-----LPAGAKISIPVFDVAVIK 287  
QY 321 FVKFPHNLNGDIHHMNIKVFNSNGLDVDPALTEGRIKE 358  
DB 288 TIQIKSGYVGNRKDTAAEAVDFFTRGLIKCPKIVGLSE 325

```

RESULT 7
QOR_HUMAN
ID QOR_HUMAN STANDARD; PRT; 329 AA.
AC Q08257;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Quinone oxidoreductase (EC 1.6.5.5) (NADPH:quinone reductase) (Zeta-
DE crystallin).
GN CRYZ.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=93221534; PubMed=846529;
RA Gonzalez P., Rao P.V., Zigler J.S. Jr.;
RT "Molecular cloning and sequencing of zeta-crystallin/quinone
RL Blochem. Biophys. Res. Commun. 191:902-907(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=94375054; PubMed=808825;
RA Gonzalez P., Rao P.V., Zigler J.S. Jr.;
RT "Organization of the human zeta-crystallin/quinone reductase gene
RT (CRYZ)".
RL Genomics 21:317-324 (1994).
CC -1- FUNCTION: DOES NOT HAVE ALCOHOL DEHYDROGENASE ACTIVITY. BINDS
CC NADP AND ACTS THROUGH A ONE-ELECTRON TRANSFER PROCESS.
CC ORTHOQUINONES ARE THE BEST SUBSTRATES. MAY ACT IN THE
CC DETOXIFICATION OF XENOBIOTICS.
CC -1- CATALYTIC ACTIVITY: NADPH + quinone = NADP(+) + semiquinone.
CC -1- SUBUNIT: Homotetramer.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- TISSUE SPECIFICITY: ONLY VERY LOW AMOUNTS IN THE LENS.
CC -1- SIMILARITY: Belongs to the zinc-containing alcohol dehydrogenase
CC family. Quinone oxidoreductase subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L13278; AAA36536.1; -.
DR EMBL; L31526; AAK40311.1; -.
DR EMBL; L31521; AAK40311.1; JOINED.
DR EMBL; L31522; AAK40311.1; JOINED.
DR EMBL; L31523; AAK40311.1; JOINED.
DR EMBL; L31524; AAK40311.1; JOINED.
DR EMBL; L31525; AAK40311.1; JOINED.
DR PIR; P04448; P04448.
DR HSP; P28304; IQOR.
DR Genew; HGNC:2419; CRYZ.
DR MIN; 123691; -.
DR GO; GO:0007601; P:vision; TAS.
DR InterPro; IPR02085; Adh zn family.
DR Pfam; PF00107; Adh_zinc_N; 1.
DR PROSITE; PS01162; QOR_ZETA_CRYSTAL; 1.
KW Oxidoreductase; NADP; Zinc.
SQ SEQUENCE 329 AA; 35206 MW; 68C1828911486D4E CRC64;

Query Match 8.4%; Score 161; DB 1; Length 329;
Best Local Similarity 24.8%; Pred. No. 0.00018;
Matches 73; Conservative 50; Mismatches 95; Indels 76; Gaps 14;

```

## RESULT 8

```

QOR_CAVPO
ID QOR_CAVPO STANDARD; PRT; 329 AA.
AC P11415;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Quinone oxidoreductase (EC 1.6.5.5) (NADPH:quinone reductase) (Zeta-
DE crystallin).
GN CRYZ.
OS Cavia porcellus (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.
OX NCBI_TaxID=10141;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=89378748; PubMed=2777081;
RA Rodokanaki A., Holmes R.K., Borras T.;
RT "Zeta-crystallin, a novel protein from the guinea pig lens is related
RT to alcohol dehydrogenases."
RL Gene 78:215-224 (1989).
RN [2]
RP CHARACTERIZATION.
RC TISSUE=Liver;
RX MEDLINE=93041895; PubMed=1420281;
RA Rao P.V., Zigler J.S. Jr.;
RT "Purification and characterization of zeta-crystallin/quinone
RT reductase from guinea pig liver."
RL Biochim. Biophys. Acta 1117:315-320(1992).
RN [3]
RP CHARACTERIZATION.
RC TISSUE=Liver;
RX MEDLINE=92112732; PubMed=1370456;
RA Rao P.V., Krishna C.M., Zigler J.S. Jr.;
RT "Identification and characterization of the enzymatic activity of
RT zeta-crystallin from guinea pig lens. A novel NADPH:quinone
RT oxidoreductase."
RL J. Biol. Chem. 267:96-102(1992).
CC -1- FUNCTION: DOES NOT HAVE ALCOHOL DEHYDROGENASE ACTIVITY. BINDS
CC NADP AND ACTS THROUGH A ONE-ELECTRON TRANSFER PROCESS.
CC ORTHOQUINONES ARE THE BEST SUBSTRATES. MAY ACT IN THE
CC DETOXIFICATION OF XENOBIOTICS.
CC -1- CATALYTIC ACTIVITY: NADPH + quinone = NADP(+) + semiquinone.
CC -1- SUBUNIT: Homotetramer.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- TISSUE SPECIFICITY: IN GUINEA PIGS IT CONSTITUTES ABOUT 10% OF THE
CC WATER SOLUBLE PROTEINS OF THE LENS.
CC -1- DISEASE: A GENOMIC MUTATION CAUSING THE DELETION OF 34 AMINO ACIDS
CC WAS CLEARLY ASSOCIATED WITH A HEREDITARY NUCLEAR CATARACT IN A
CC LINE OF STRAIN 13 GUINEA PIGS.
CC -1- SIMILARITY: Belongs to the zinc-containing alcohol dehydrogenase
CC family. Quinone oxidoreductase subfamily.

```

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation  
at the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see http://www.isb-sib.ch/announce/  
or send an email to licens@isb-sib.ch).  
-----  
EMBL; M26936; AAA37035.1; --  
DR PIR; JS0230; CYGPZ.  
DR HSPP; P28304; IQOR.  
DR InterPro; IPR002085; Adh\_zn\_family.  
DR InterPro; IPR002364; QOR\_zeta\_crystal.  
DR Pfam; PF001107; ADH\_zinc\_N; 1.  
DR PROSITE; PS01162; QOR\_ZETA\_CRYSTAL; 1.  
KW Oxidoreductase; NADP; Zinc; Eye lens protein.  
SQ SEQUENCE 329 AA; 3520 MW; 1463632C672C234F CRC64;  
-----  
Query Match 8.3%; Score 159; DB 1; Length 329;  
Best Local Similarity 24.2%; Pred. No. 0.00025;  
Matches 71; Conservative 50; Mismatches 96; Indels 76; Gaps 14;  
-----  
QY 19 VKTDVSVPELKEGTALVKVEAVAGNPDTWKHIAYKIGPEGSIL----GCDIAGTVVKLG 74  
|::||::||::||::||::||::||::||::||::||::||::||:  
Db 24 VQSDVAVPVPKDHQVLIKVHACGINPVE--TYIRSGETRIPLLPYTPGTGVAGVESIGN 82  
|::||::||::||::||::||::||::||::||::||::||::||:  
QY 75 NASDTDLKVGDGTGFVFHGASQDTPKNCAFAEYARVYPPLFKSNLTHTADEISEGPVK- 133  
|::||::||::||::||::||::||::||::||::||::||::||:  
Db 83 DVSA-FKKGRVFP-----TTSTISGYAEYALASDHTVYRL-----PEKL 121  
|::||::||::||::||::||::||::||::||::||::||::||:  
QY 134 NFESAASLPVSLTTAGVSLCHLGSKMEWHPSTPQHTHLLIMGWATVGQQLIQAKHI 193  
|::||::||::||::||::||::||::||::||::||::||::||:  
Db 122 DFRQGAIGIPYFTA---CRAL-----FHSAKAGESVLVHGASGVGLAACQIAR-- 170  
|::||::||::||::||::||::||::||::||::||::||::||:  
QY 194 NAY-TKIV-TVASKKHKLKSYGADVDFDYHDAGVTIEQIKSPYNLOHVIDAVGSSESI 251  
|::||::||::||::||::||::||::||::||::||::||::||:  
Db 171 -AYGLKVLGTAGTEGGQVKVQLNGAHEVFNRDAHYIDEIKK-----SIGEGK-- 217  
|::||::||::||::||::||::||::||::||::||::||::||:  
QY 252 PEAYKYTADSLPATLLEVPMTIESPEETRKONKVIDITLLYRASQEILLG 304  
|::||::||::||::||::||::||::||::||::||::||::||:  
Db 218 -----VDVIIEMLANVNLSNDLKL-SCGGRVIIVG 247  
|::||::||::||::||::||::||::||::||::||::||::||:  
-----  
RESULT 9  
QOR\_MOUSE STANDARD; PRG 331 AA.  
ID QOR\_MOUSE AC F47199; Q62508; Q99L63;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Quinone oxidoreductase (EC 1.6.5.5) (NADPH:quinone reductase) (Zeta-  
DE DE crystallin).  
GN GN Mus musculus (Mouse).  
OS OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OC NCBI\_TaxID=10090;  
RN RN [1]  
RP RP SEQUENCE FROM N.A.  
RC TISSUE=Liver;  
RX MEDLINE=94224126; PubMed=8170370;  
RA Gonzalez P., Hernandez-Calzadilla C., Rao P.V., Rodriguez I.R.,  
RA Zigler J.S. Jr., Borras T.;  
RT "Comparative analysis of the zeta-crystallin/quinnone reductase gene  
RT in guinea pig and mouse."  
RL Mol. Biol. Evol. 11:305-315(1994).  
RN [2]  
RP RP SEQUENCE FROM N.A.  
RC TISSUE=Breast tumor;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA [2]

```
Db 120 -TLNFRGGAALGIPYFTA-----CRAL-----FHSARARAGESVLVHGASGVGLTACQIA 169
Qy 191 KHINAYTKIVTASKEKELKSKGADVDYHDAGVIEQIKSKYPNLQHVDAVGSSEDS 250
Db 170 R-ANGLKVLGTAGSEBEGKVLQNGAENFVNKKAINDIKM-----SVGDKDK 218
Qy 251 IPEAYKVTADSLPATLLEVPMTTIESIPBEIRKONKVIDITLLYRAGQBIILLGATRPFA 310
Db 219 -----GVDVITIELANENLNDLKL--SHGGRVVVVGCRGPI 254
Qy 311 SPEYHEATVKPVKEINPHLNN 331
Db 255 EINPRDTWAKETSIIIGVSLSS 275

RESULT 10
QD_LEIAM
ID_QD_LEIAM STANDARD; PRT; 340 AA.
AC P42865;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Possible quinone oxidoreductase (EC 1.6.5.3) (NADPH:quinone
DE reductase) [P36].
OS Leishmania amazonensis.
OC Eukaryota; Eulenzozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5659;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95107343; PubMed=7808470;
RA Liu X., Chang K.P.;
RT "Identification by extrachromosomal amplification and overexpression
RT of a zeta-crystallin/NADPH-oxidoreductase homologue constitutively
RT expressed in Leishmania spp.";
RL Mol. Biochem. Parasitol. 66:201-210(1994).
CC -1- CATALYTIC ACTIVITY: NADPH + quinone = NADP(+) + semiquinone.
CC -1- SIMILARITY: Belongs to the zinc-containing alcohol dehydrogenase
CC family. Quinone oxidoreductase subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L11705; AAA73554.1; -
DR InterPro; IPR002085; Adh zn family.
DR InterPro; IPR002364; QOR_zeta_crystal.
DR Pfam; PF00107; ADH_zinc_N; 1
DR PROSITE; PS01162; QOR_ZETA_CRYSTAL; 1.
KW Oxidoreductase; NADP; Zinc.
SQ SEQUENCE 340 AA; 36272 MW; 4D9255AB568566235 CRC64;

Query Match
Best Local Similarity 8.1%; Score 155; DB 1; Length 340;
Matches 92; Conservative 58; Mismatches 151; Indels 90; Gaps 19;

Qy 1 MSVPTTKAVIIEGDKAVKVTDSV-----PELKEGTALVKVEAVAGNPTDWRKIAYKI 54
Db 1 MSSPSNFKQLQVSLSKDFRSSTSVVEAHLPEEVEPGNVRVSVKYAGVNASDLNF----- 55
Qy 55 GPEGSL-----GCDINGTVKVLGPNASTDLKVGDTGFGFVHGASQTDPKNGAPAE 105
Db 56 -TNGSYFNQVQPPFDPCGFEAGTVQIGAGVA-NVKVGD-----HVVLM---QYGCFAE 104
Qy 106 YARVVPPLFYKSLNLTHTSADRISEG--PVKNFESAAS-LPVSLTAGVSLCHLGSKMWE 162
Db 105 F-----LDAPERCIPVPELKPESVLPVPSAULTAVAL-GEVGRVKKG 146
Qy 163 HPSTPQHTHPLLWGGATAVGQQLIQVAKHINAYTKIVTASKEKELKSKGADVDVFDY 222
```

```
Db 147 DVA-----LVTAAAGGTGQIAVQLLKHVYCTVIGTCSSEKAFLKSGICDHVINY 198
Qy 223 HDAGVIEQIKSKYP-NLQHVDAVGSSEDSIPYAYKVYADSLPATL-----EVVPM 272
Db 199 KTESLDRHLHELCPKGVVYECVGGH-TFNDARVHVAVHARVVIIIGSTSSYKSGEVVPP 257
Qy 273 TIESIPBEIRKONKVIDITLLYR-ASGQBIILLGATRPFAPEYHEATVKPVKEINPHLNN 331
Db 258 SDPS-----GTSVTMLLLVKASLNGFFL-----POPHDVI PKYMANLLQYLKA 301
Qy 332 GDTH-HMNIKVFNSGLDDVPALTEGKEGKN 361
Db 302 GQVKLFVDKKVF-HGLSSVADAVDHLVSGAN 331

RESULT 11
YM27_YEAST
ID_YM27_YEAST STANDARD; PRT; 365 AA.
AC P28635;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Hypothetical 41.6 kDa protein in IMP1-HLJ1 intergenic region (RF1095).
DE YMR152W OR YMR375.22 OR YMR520.01.
GN OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91360060; PubMed=1886606;
RA Behrens M., Michaelis G., Pratje E.;
RT "Mitochondrial inner membrane protease 1 of Saccharomyces cerevisiae
RT shows sequence similarity to the Escherichia coli leader peptidase.";
RL Mol. Gen. Genet. 228:167-176(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=S288c / AB972;
RX PubMed=9169872;
RA Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T.,
RA Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.,
RA Jagels K., Lyne G., Moule S., Odell C., Pearson D., Rajandream M.A.,
RA Rice P., Skelton J., Walsh S., Whitehead S., Barrell B.G.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome
RT XIII.";
RL Nature 387:90-93(1997).
CC -1- SIMILARITY: SOME, TO YEAST AST1/AST2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; S55518; AAB19702.1; -
DR EMBL; Z47071; CAA87367.1; -
DR EMBL; Z49705; CAA89788.1; -
DR PIR; S50409; S50409.
DR SGD; S0004760; YIM1.
DR GO; GO:0005811; C:lipid particle; IDA.
DR GO; GO:0005743; C:mitochondrial inner membrane; IDA.
DR GO; GO:0008233; F:peptidase activity; IMP.
DR GO; GO:0006627; P:mitochondrial processing; IMP.
DR InterPro; IPR002085; Adh zn family.
DR Pfam; PF00107; ADH_zinc_N; 1.
KW Hypothetical protein; Transmembrane
TRANSMEM 140 158
FT CONFLICT 121 121 V -> I (IN REF. 1).
FT CONFLICT 234 234 S -> I (IN REF. 1).
FT CONFLICT 280 280 D -> N (IN REF. 1).
```

```
FT CONFLICT 296 296 L -> S (IN REF. 1).
SQ SEQUENCE 365 AA; 41637 MW; 88F6453D9E918A16 CRC64;

Query Match
Best Local Similarity 8.0%; Score 154; DB 1; Length 365;
Matches 63; Conservative 49; Mismatches 116; Indels 40; Gaps 12;

QY 6 TOKAVILGDKAVKVTDSVPPEL-----EGTALVKVAVAGNPTDW-KH---IAYKIGPE 57
DB 7 TNKSVYVNTPTVITSSSELDLSCVQDDVEVIEVHAALNPIDFTHQLCNSYIFGKY 66
QY 58 GSILGCDIAGTVVVLGPNASTDLKVGDTGFGF---VHGASQDTPKNGAFAYARVYP--- 111
DB 67 PKTYSRDSYGVIIKAGKDQVNRKVGDKVNGWYSHVIG-----ERGLTHYLILNPAKD 120
QY 112 -PLPKSNLTHSTADEISEGPKVFESNASLPVSLTTAGVSLCHLGSKMEHWPSTPOHT 170
DB 121 VPI---THMVRVPKDE---NDPVDYFVAAAAPLFTGTAFTSL---YDFKDKDWTSDS--- 168
QY 171 HPLILWGATAVGQQLIQVAKHINAYTKIVTVASKKHEKLKSKYGADVDYHDAGVIEQ 230
DB 169 -KVLVIGASTSVSAFVHIAKNYFNIGTVGICNSIERNKGLGYDYLVPYDEGSIVEN 227
QY 231 IKS-KYPNLOH-----VIDAVGSEDSIP 252
DB 228 VKLKQSVLENDKDFMIFDSVGNHDFRP 255

RESULT 12
P1 ARATH
ID P1 ARATH STANDARD; PRT; 345 AA.
AC Q35172;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable NADP-dependent oxidoreductase P1 (EC 1.1.1.1).
GN P1 OR AT5G16970 OR P2K13 120.
OS Arabidopsis thaliana (Mouset-ear crease).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1];
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=96064691; PubMed=7592828;
RA Babylchuk E., Kushnir S., Belles-Boix E., van Montagu M., Inze D.;
RT "Arabidopsis thaliana NADPH oxidoreductase homologs confer tolerance
of yeasts toward the thiol-oxidizing drug diamide.";
RL J. Biol. Chem. 270:26224-26231(1995).
RN [2];
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=21016721; PubMed=11130714;
RA Tabata S., Kaneko T., Nakamura Y., Kotani H., Kato T., Asamizu E.,
RA Miyajima N., Saeamoto S., Kimura T., Hosouchi T., Kawashima K.,
RA Kohara M., Matsumoto M., Matsuno A., Muraki A., Nakayama S.,
RA Watanabe N., Naruo K., Okumura S., Shinpo S., Takeuchi C., Wada T.,
RA Nakazaki N., Yanada M., Yasuda M., Sato S., de la Bastide M.,
RA Huang E., Spiegel L., Gnoj L., O'Shaughnessy A., Preston R.,
RA Habermann K., Murray J., Johnson D., Rohlfing T., Nelson J.,
RA Stoneking T., Pepin K., Spieth J., Sekhon M., Armetrong J., Becker M.,
RA Belter E., Cordum H., Cordes M., Courtney L., Courtney W., Dante M.,
RA Du H., Edwards J., Fryman J., Haakensen B., Lamar E., Latrelle P.,
RA Leonard S., Meyer R., Mulvaney E., Ozerky P., Riley A., Strommatt C.,
RA Wagner-McPherson C., Wollman A., Yoakum M., Bell M., Dedhia N.,
RA Patnell L., Shah R., Rodriguez M., Hoon See L., Vil D., Baker J.,
RA Kirchoff K., Toth K., King L., Bahret A., Miller B., Marra M.,
RA Martensen R., McCombie W.R., Wilson R.K., Murphy G., Bancroft I.,
RA Volckaert G., Wambutt R., Duesterhoeft A., Stiekema W., Pohl T.,
RA Entian K.-D., Terryn N., Hartley N., Bent E., Johnson S.,
RA Langham S.-A., McCullagh B., Robben J., Grymonprez B., Zimmermann W.,
RA Ransperger U., Wedler H., Balke K., Wedler E., Peters S.,
```

Saccharomycetales; Saccharomycetaceae; Kluyveromyces.  
 NCBI\_TaxID=4911;  
 (1)  
 SEQUENCE FROM N.A.  
 STRAIN=ATCC 12424;  
 MEDLINE=93250057; PubMed=8485163;  
 Ladriere J.M., Delcour J., Vandenhaute J.;  
 "Sequence of a gene coding for a cytoplasmic alcohol dehydrogenase  
 from Kluyveromyces marxianus ATCC 12424";  
 Biochim. Biophys. Acta 1173:99-101(1993).  
 CC -1- CATALYTIC ACTIVITY: An alcohol + NAD(+) = an aldehyde or ketone +  
 CC NADH.  
 CC -1- COFACTOR: Binds 2 zinc ions per subunit (By similarity).  
 CC -1- SUBUNIT: Homotetramer.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- SIMILARITY: Belongs to the zinc-containing alcohol dehydrogenase  
 CC family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X60224; CAA42785.1; -;  
 CC PIR; S32521; S32521.  
 CC InterPro; IPR002328; ADH\_zinc.  
 CC InterPro; IPR002085; Adh\_zn\_family.  
 CC Pfam; PF00107; ADH\_zinc\_N; 1.  
 CC PROSITE; PS00059; ADH\_ZINC; 1.  
 CC Oxidoreductase; Zinc; Metal-binding; NAD; Multigene family.  
 FT METAL 44 44 ZINC 1 (CATALYTIC) (BY SIMILARITY).  
 FT METAL 67 67 ZINC 1 (CATALYTIC) (BY SIMILARITY).  
 FT METAL 98 98 ZINC 2 (BY SIMILARITY).  
 FT METAL 101 101 ZINC 2 (BY SIMILARITY).  
 FT METAL 104 104 ZINC 2 (BY SIMILARITY).  
 FT METAL 112 112 ZINC 2 (BY SIMILARITY).  
 FT METAL 154 154 ZINC 1 (CATALYTIC) (BY SIMILARITY).  
 FT METAL 155 155 ZINC 1 (CATALYTIC) (BY SIMILARITY).  
 CC SEQUENCE 348 AA; 37158 MW; A75D2EBE82E355BD CRC64;  
 CC -----  
 CC Query Match 8.0%; Score 153.5; DB 1; Length 348;  
 CC Best Local Similarity 22.6%; Pred. No. 0.00067;  
 CC Matches 90; Conservative 53; Mismatches 142; Indels 113; Gaps 20;  
 CC -----  
 CC 1 MSVPTTKAVII--EGDKAVVKTDSVPPELKEGTALVKVE-----AVAGN-PTDW 47  
 CC 1 MAIPETQKGVIFYEHGGLQYK-DIPVKKPNELLINVKYSGVCHTDLHAWQGDWPLDT 59  
 CC -----  
 CC 48 KHIAVKIGPEGSILGCDIAGTVVKGPNASTDLKVG-D-TGFGFVHGASQT-----96  
 CC 60 K-LPLVGGHEG-----AGIVVANGENV-TGWEIGDVAGIKWLNGSCMCECELNEP 110  
 CC -----  
 CC 97 -DPK-----NGAFAYARYVPPLFYKSLNTHSTADEISEGPV-KNFESAASLPVSLT 146  
 CC 111 NCPRADLSGYTHDGSFQYA-----TADAVQAARIPKNVDLAEVAPI--L 153  
 CC -----  
 CC 147 TAGVSLCHLGSKM-----EWPSTPQTHPLLWGATAVGQQILQVAKHINATKIVTV 202  
 CC 154 CAGVTYVYKALKSAHIKAGDW-----VAISGACGLGLSLAIQYAKAM-GYRVLGID 202  
 CC -----  
 CC 203 ASKKEKLLKSGYADDVFDYHDAGVIEQIKSPNLOHVIDAVGSEDSIPKAYKVTADSL 262  
 CC 203 AGDEKAKLFKEGLG-----EYFDFTKKNVAEIVATNGVA 240  
 CC -----  
 CC 263 PATLLEVVPMTIESIPBEIRKDNVKIDITLY-RASGOEILLGATR-PPASPEYHEATVK 320  
 CC 241 HAVI-----NVSVSEAAISTSVLYTRNGTVLVGLPRDAQCKSDVFNQVVK 287  
 CC -----  
 CC 321 FVKFINPHLNGDIIHNNIKVFSNGLDDVPALTKGIKE 358  
 CC 288 SISIVGSYVGNRADTREALDFFSRGLVKAPIKILGLSE 325

RESULT 14  
 ADH2\_PICST STANDARD; PRT; 348 AA.  
 AC O13309; O00090;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Alcohol dehydrogenase II (EC 1.1.1.1) (ADH 1).  
 GN ADH2 OR ADH1.  
 OS Pichia stipitis (Yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Pichia.  
 OX NCBI\_TaxID=4924;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 58785 / CBS 6054;  
 RX MEDLINE=98207839; PubMed=9546172;  
 RA Cho J.Y., Jeffries T.W.;  
 RT "Pichia stipitis genes for alcohol dehydrogenase with fermentative  
 RT and respiratory functions";  
 RL Appl. Environ. Microbiol. 64:1350-1358(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CBS 5774;  
 RX MEDLINE=99019018; PubMed=9802210;  
 RA Pascoth V., Schaefer B., Liebel B., Weierstall T., Kliner U.;  
 RT "Molecular cloning of alcohol dehydrogenase genes of the yeast Pichia  
 RT stipitis and identification of the fermentative ADH.";  
 RL Yeast 14:1311-1325(1998).  
 CC -1- FUNCTION: CONVERTS ETHANOL TO ACETALDEHYDE.  
 CC -1- CATALYTIC ACTIVITY: An alcohol + NAD(+) = an aldehyde or ketone +  
 CC NADH.  
 CC -1- COFACTOR: Binds 2 zinc ions per subunit (By similarity).  
 CC -1- SUBUNIT: Homotetramer (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- SIMILARITY: Belongs to the zinc-containing alcohol dehydrogenase  
 CC family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AF008244; AAC49990.1; -;  
 CC EMBL; Y13238; CAA73690.1; -;  
 CC InterPro; IPR002328; ADH\_zinc.  
 CC InterPro; IPR002085; Adh\_zn\_family.  
 CC Pfam; PF00107; ADH\_zinc\_N; 1.  
 CC PROSITE; PS00059; ADH\_ZINC; 1.  
 CC Oxidoreductase; Zinc; Metal-binding; NAD; Multigene family.  
 FT METAL 44 44 ZINC 1 (CATALYTIC) (BY SIMILARITY).  
 FT METAL 67 67 ZINC 1 (CATALYTIC) (BY SIMILARITY).  
 FT METAL 98 98 ZINC 2 (BY SIMILARITY).  
 FT METAL 101 101 ZINC 2 (BY SIMILARITY).  
 FT METAL 104 104 ZINC 2 (BY SIMILARITY).  
 FT METAL 112 112 ZINC 2 (BY SIMILARITY).  
 FT METAL 154 154 ZINC 1 (CATALYTIC) (BY SIMILARITY).  
 FT METAL 155 155 ZINC 1 (CATALYTIC) (BY SIMILARITY).  
 CC SEQUENCE 348 AA; 35565 MW; F6C813C98F56A148 CRC64;  
 CC -----  
 CC Query Match 7.9%; Score 152.5; DB 1; Length 348;  
 CC Best Local Similarity 24.3%; Pred. No. 0.00079;  
 CC Matches 97; Conservative 51; Mismatches 136; Indels 115; Gaps 21;  
 CC -----  
 CC 1 MSVPTTKAVIIE--GDKAVVKTDSVPPELKEGTALVKVEAVAGNPTD---WK-----48  
 CC 1 MSVPTTKAVIFETNGGPLYK-DIPVKKPNELLINVKYSGVCHTDLHAWQGDWPLDT 59





```

CC TO SINGLE-STRANDED DNA.
CC -1- SUBUNIT: Homotetramer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to the zinc-containing alcohol dehydrogenase
CC family. Quinone oxidoreductase subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
CC -----
CC EMBL: U70048; RAD10290.1; -.
CC DR HSSP: P28304; IQOR.
CC DR InterPro: IPR002085; Adh_zn_family.
CC DR Pfam: PF00107; ADH_zinc_N; 1.
CC DR PROSITE: PS01162; QOR_ZETA_CRYSTAL; 1.
CC KW Zinc; DNA-binding.
CC SQ SEQUENCE 330 AA; BD24ACD857BA8673 CRC64;

Query Match 7.78; Score 148; DB 1; Length 330;
Best Local Similarity 24.0%; Pred. No. 0.0016;
Matches 71; Conservative 54; Mismatches 91; Indels 80; Gaps 15;

Qy 19 VKTDSVPELKEGTALVKVEAVAGNPTDM--KHAYKIGPEGSTL---GCDIAGTVVKL 72
Db 24 LQSDVAVPIPEKHQVLIKQACGVNPDVTYRSQTHNKP---LLPTGDFVAGIIEAV 80
Qy 73 GPNASTDLKVGDTGFGFVHGASQTDPKNGAFAYRVPYPLFYKSNLTHSTADEISEGPV 132
Db 81 GESVSA-FKKGDRVF-----TTRTISGGVAYEALAAADHTVYL-----PE 119
Qy 133 K-NFESAASLPVSLTTAGVSLCHHLGSKMEWHFSTPQHTPLLIWGATATVGOQLIQAK 191
Db 120 KLDFKQGAALIGPYFTAYRALLYSA-----PVKPGES--VLVHGASGGVGAACQIAR 170
Qy 192 HINAY-TKIV-TVASKKHKLKSYGADDVFDVHDAGVYEQIKSKYPNLQHVDAVGSSED 249
Db 171 ---AYGLKVGITSTEGQKIVLENGAKVFNHKEANYIDKIKK-----SVGEKG 217
Qy 250 SIPEAYKVTDASLPATLLEVPVMTIESIPEIRKDNVKIDITLLYRASGOEILLGA 305
Db 218 -----VDVIEMLANVNLNDLNLSSH-GGRVIVVGS 248

RESULT 17
QOR_LAMGU STANDARD; PRT; 330 AA.
AC Q28452;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DE Quinone oxidoreductase (EC 1.6.5.5) (NADPH:quinone reductase) (Zeta-
DE crystallin).
GN CRYZ.
OS Lama guanicoe (Guanaco).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Tylopoda; Camelidae; Lama.
OX NCBI_TaxID=9840;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lens;
RX MEDLINE=96073447; PubMed=7476124;
RA Gonzalez P., Rao P.V., Nunez S.B., Zigler J.S. Jr.;
RT "Evidence for independent recruitment of zeta-crystallin/quinone
RT reductase (CRYZ) as a crystallin in camels and hystriomorph
RT rodents."
RL Mol. Biol. Evol. 12:773-781(1995).
CC -1- FUNCTION: DOES NOT HAVE ALCOHOL DEHYDROGENASE ACTIVITY. BINDS
CC NADP AND ACTS THROUGH A ONE-ELECTRON TRANSFER PROCESS.

```

```

CC ORTHOQUINONES ARE THE BEST SUBSTRATES. MAY ACT IN THE
CC DETOXIFICATION OF XENOBIOTICS (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: NADPH + quinone = NADP(+) + semiquinone.
CC -1- SUBUNIT: Homotetramer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to the zinc-containing alcohol dehydrogenase
CC family. Quinone oxidoreductase subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
CC -----
CC EMBL: L34159; AAA99986.1; -.
CC DR HSSP: P28304; IQOR.
CC DR InterPro: IPR002085; Adh_zn_family.
CC DR Pfam: PF00107; ADH_zinc_N; 1.
CC DR PROSITE: PS01162; QOR_ZETA_CRYSTAL; 1.
CC KW Oxidoreductase; NADP; Zinc.
CC SQ SEQUENCE 330 AA; 35187 MW; E784E414D2BA23D6 CRC64;

Query Match 7.7%; Score 148; DB 1; Length 330;
Best Local Similarity 23.1%; Pred. No. 0.0016;
Matches 67; Conservative 51; Mismatches 104; Indels 68; Gaps 11;

Qy 19 VKTDSVPELKEGTALVKVEAVAGNPTDM--KHAYKIGPEGSTL---GCDIAGTVVKLGN 75
Db 24 LQSDVAVPIPEKHQVLIKQACGVNPDVTYRSQTHNKP---LLPTGDFVAGIIEAV 83
Qy 76 ASTDLKVGDTGFGFVHGASQTDPKNGAFAYRVPYPLFYKSNLTHSTADEISEGPVKNF 135
Db 84 VSA-FKKGDRVF-----TTRTISGGVAYEALAAADHTVYL-----LPGL 123
Qy 136 ESAASLPVSLTTAGVSLCHHLGSKMEWHFSTPQHTPLLIWGATATVGOQLIQAKHINA 195
Db 124 QKGAIGVPTAYRALLYSACA-----GESVLHGASGGVGAACQIARAC-C 173
Qy 196 YTKIVTVASKKHKLKSYGADDVFDVHDAGVYEQIKSKYPNLQHVDAVGSSEDSPEAY 255
Db 174 FKVLGTAGTEGQRVVLQNGAHEVFNHREDINIDKIKK-----SVGEKG----- 217
Qy 256 KVTADSLPATLLEVPVMTIESIPEIRKDNVKIDITLLYRASGOEILLGA 305
Db 218 -----IDVIEMLANVNLNDLNLSSH-GGRVIVVGS 248

RESULT 18
P2_ARATH STANDARD; PRT; 343 AA.
AC Q39173; Q9LFPK3;
DT 01-NOV-1997 (Rel. 35, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Quinone oxidoreductase (EC 1.6.5.5) (NADPH:quinone reductase) (Zeta-
DE crystallin).
GN P2 OR AT5G16990 OR F2K13_140.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsia.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=96064691; PubMed=7592828;
RA Babychuk E., Kuehnir S., Belles-Boix E., van Montagu M., Inze D.;
RT "Arabidopsis thaliana NADPH oxidoreductase homologs confer tolerance
RT of yeasts toward the thiol-oxidizing drug diamide."
RL J. Biol. Chem. 270:26224-26231(1995).
RN [2]

```

```

SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=21016721; PubMed=11130714;
RA Tabata S., Kaneko T., Nakamura Y., Kotani H., Kato T., Asamizu E.,
RA Miyajima N., Sasamoto S., Kimura T., Hosouchi T., Kawashima K.,
RA Kohara M., Matsumoto M., Matsumoto A., Muraki A., Nakayama S.,
RA Nakazaki N., Naruo K., Okumura S., Shino S., Takeuchi C., Wada T.,
RA Watanabe A., Yanada M., Yasuda M., Sato S., de la Bastide M.,
RA Huang E., Spiegel L., Gnoj L., O'Shaughnessy A., Preston R.,
RA Habermann K., Murray J., Johnson D., Rohlfing T., Nelson J.,
RA Stoneking T., Pepin K., Spieth J., Sekhon M., Armstrong J., Becker M.,
RA Belter E., Cordum H., Cordes M., Courtney L., Courtney W., Dante M.,
RA Du H., Edwards J., Fryman J., Haakensen B., Lamar E., Latreille P.,
RA Leonard S., Meyer R., Mullaney E., Ozersky P., Riley A., Strommatt C.,
RA Wagner-McPherson C., Wollam A., Yeakum M., Bell M., Dedhia N.,
RA Parnell L., Shah R., Rodriguez M., Hoon See L., Vil D., Baker J.,
RA Kirchoff K., Toth K., King L., Bahret A., Miller B., Marra M.,
RA Martensen R., McCombie W.R., Wilson R.K., Murphy G., Bancroft I.,
RA Volckaert G., Wambutt R., Duesterhoeft A., Stiekema W., Pohl T.,
RA Entian K.-D., Terryn N., Hartley N., Bent E., Johnson S.,
RA Langham S.-A., McCullagh B., Robben J., Grymonprez B., Zimmermann W.,
RA Ramsperger U., Wedler H., Baake K., Wedler E., Peters S.,
RA van Staveren M., Dirkse W., Mooijman P., Klein Lankhorst R.,
RA Weitzenecker T., Bothe G., Rose M., Hauf J., Berneiser S., Hempel S.,
RA Feldpausch M., Lamberth S., Villarroel R., Gielen J., Ardiles W.,
RA Bents O., Lemcke K., Kolesov G., Mayer K.P.X., Rudd S., Schoof H.,
RA Schueller C., Zaccaria P., Mewes H.-W., Bevan M., Franz P.F.;
RT "Sequence and analysis of chromosome 5 of the plant Arabidopsis
RT thaliana.";
RL Nature 408:823-826(2000).
RN [3]
SEQUENCE FROM N.A.
RP STRAIN=cv. Columbia;
RC Shinzaki K., Davis R.W., Ecker J.R., Theologis A.;
RA "RIKEN Arabidopsis full length cDNA clones (RAFLs) sequenced by the
RT SSP consortium (Salk/Stanford/PGSC).";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
CC -I- FUNCTION: MAY PLAY A DISTINCT ROLE IN PLANT ANTIOXIDANT DEFENSE
CC AND IS POSSIBLY INVOLVED IN NAD(P)/NAD(P)H HOMEOSTASIS.
CC -I- SIMILARITY: BELONGS TO THE L4BD FAMILY OF NADP-DEPENDENT
CC OXIDOREDUCTASE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z49268; CAA8262.1; --
DR EMBL; AL391141; CAC01712.1; --
DR EMBL; AY091320; AAM14259.1; --
DR EMBL; AY065253; AAL38729.1; --
DR PIR; S57612; S57612.
DR PIR; T51554; T51554.
DR InterPro; IPR002085; Adh_zn_family.
DR Pfam; PF00107; Adh_zn_N; 1.
KW Oxidoreductase; NADP.
FT CONFLICT 15 15 MISSING (IN REF. 1).
SQ SEQUENCE 343 AA; 37989 MW; 491698EF58BA82DA CRC64;

Query Match 7.6%; Score 146; DB 1; Length 343;
Best Local Similarity 21.6%; Pred. No. 0.0023;
Matches 86; Conservative 61; Mismatches 160; Indels 92; Gaps 18;

QY 5 TTOKAVII-----EGDKAVVKTDVS--VPELKEGTALVK-----VEAVAGNP 44
DB 2 TTNKQVLPKHVSQFPKESDFNTTTVELRVPEGSK-SVLVKNLYLSCDPYWRSMGKP 60
QY 45 TDWKHIAHYKIGPEGLGCDIAGTVVKLGNASTDLKVGDTGFGFVHGASQTDPKNGAFA 104
DB 61 DPSSALAAQAPAGPKPIYGYGVSRVIESGHP-----DYKKGLLWGV-----GWE 105

SEQUENCE FROM N.A.
105 EVARVVPPLFYKSNLTHSTADEISEGVKVFSAASLPVSLTTAGVSLCHLHLSKMEWHP 164
106 EKSVTIPMAHMEFKIOH-----TDVPLSYITGLLMPGWTAYAGF-----YEV 148
165 STPOQTHPLLWGATAVGQQOIQIAKHINAYTKIVTVASKKHEKLLKS-YGADDVFDY- 222
149 CSPKKGTEVTVSAASGAVGQVQPAKMMGCVV-VGSAGSKEKVDLLTKFGFDDAFNYK 207
223 HDAGVIEQIKSKYP-NLQHVDAVGSSEDSIPRAYKVTADSLPATLLEVVPMTIESIPEEI 281
208 EESDLSAALKRCRCPKIDMYFENVGK-----MLDAVLNMMPHGRGIAVCGMI 255
282 RKNVKNIDITLVRASGQS-----ILLGATRFP--ASPEYHEATVKVKFINPHLNG 332
256 SQ-----YNLENQSGVHNLSNIIYKRIQGVVADVFYDKYPKLELVLPRKEG 305
333 DIHNMNKFVNSGLDDVPALTEGIRKGN--KNVKYVAR 369
306 KITV--VEDVADGLEKAPALVGLPHGKNGVQVVVIAR 342

RESULT 19
ADH1_KLJULA
ID ADH1_KLJULA STANDARD; PRT; 350 AA.
AC P20369;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Alcohol dehydrogenase I (EC 1.1.1.1).
GN ADH1.
OS Kluyveromyces lactis (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetaceae; Saccharomycetaceae; Kluyveromyces.
OX NCBI_TaxID:28985;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90273773; PubMed=2190430;
RA Saliola M., Shuster J.R., Falcone C.;
RT "The alcohol dehydrogenase system in the yeast, Kluyveromyces
RT lactis.";
RL Yeast 6:193-204(1990).
CC -I- CATALYTIC ACTIVITY: An alcohol + NAD(+) = an aldehyde or ketone +
CC NADH.
CC -I- COPACTOR: Binds 2 zinc ions per subunit.
CC -I- SUBUNIT: Homotetramer.
CC -I- SUBCELLULAR LOCATION: Cytoplasmic.
CC -I- SIMILARITY: Belongs to the zinc-containing alcohol dehydrogenase
CC family.
DR PIR; S09475; S09475.
DR InterPro; IPR002328; Adh_zinc.
DR InterPro; IPR002085; Adh_zn_family.
DR Pfam; PF00107; Adh_zn_N; 1.
DR PROSITE; PS00059; ADH_ZINC; 1.
KW Oxidoreductase; Zinc; Metal-binding; NAD; Multigene family.
FT METAL 46 46 ZINC 1 (CATALYTIC).
FT METAL 69 69 ZINC 1 (CATALYTIC).
FT METAL 100 100 ZINC 2.
FT METAL 103 103 ZINC 2.
FT METAL 106 106 ZINC 2.
FT METAL 114 114 ZINC 2.
FT METAL 156 156 ZINC 1 (CATALYTIC).
SQ SEQUENCE 350 AA; 37260 MW; 3D71BE2D5CC86119 CRC64;

Query Match 7.6%; Score 145.5; DB 1; Length 350;
Best Local Similarity 23.0%; Pred. No. 0.0025;
Matches 91; Conservative 49; Mismatches 146; Indels 109; Gaps 18;

QY 2 SUPTTOKAVII-EGDKAVVKTDVSUPELKEGTALVKVEAVAGNPFD---WK---HAYKI 54
DB 4 SIPETQKQVIFYENGGELOQKDFIPVKPRANELLINVKYSGVCHTDHAWKGDWPLPTKL 63
QY 55 ----GPEGSILGCDIAGTVVKLGNASTDLKVGDTGFGFVHGASQTDPKNGAFAEYARV 109

```





QY 106 YARVPPPLFYKSNLTHSTADISE--GPVKNFESAASLPVSLTTAGVSLCHHLSKMEWH 163  
Db 124 YLVI--PAFNAFKIPDNISDDLAFIDP---FGNAVHTALSFDLVGED----- 166  
QY 164 PSTQHTPLLIWGTATVAGQOLQVAKHINAYTKIVTVASKHEKLLKSYGADDFDY- 222  
Db 167 -----VLVSGAGPIGIMAAVAKHVGARNVIT-----DVNEYR 200  
QY 223 ----HDAGVIEQIKSKYPNLQHVDAVGSSESI-----PRAYKVTADSL-----PAT 265  
Db 201 LELARKMGITRAVNAVAKENLNDVMTLGTGEGFDVGLMSGAPPAFTMLDTMNHGGRIA 260  
QY 266 LLEVPMWTIESIPEIRKONKVIDIT-LLYRASQOEILLGLATRPASPVEYHEATVKVFK 324  
Db 261 MLGIPP-----SDMSIDWTKVIFKGLFTKIYGREMF-----ETWYKMAAL 301  
QY 325 INPHLNGDI--HNMNIKVPFSGLDVDPALTEGIEKKG 360  
Db 302 IQSGLDLSPLIITRPSIDDFQKGF-----AMRSG-QSGK 335  
RESULT 23  
ADH1\_BACST STANDARD; PRT; 337 AA.  
AC Fl2311;  
DT 01-OCT-1989 (Rel. 12, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Alcohol dehydrogenase [EC 1.1.1.1] (ADH-T).  
GN ADH1.  
OS Bacillus stearothermophilus.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Geobacillus.  
OX NCBI\_TaxID=1422;  
RN [1]  
RP SEQUENCE FROM N.A., AND MUTAGENESIS.  
RC STRAIN=NCA 1503; PubMed=1735726;  
RX MEDLINE=92138636; PubMed=1735726;  
RA Sakoda H., Inanaka T.;  
RT "Cloning and sequencing of the gene coding for alcohol dehydrogenase  
of Bacillus stearothermophilus and rational shift of the optimum  
pH";  
RL J. Bacteriol. 174:1397-1402(1992).  
RN [2]  
RP SEQUENCE OF 1-45.  
RX MEDLINE=7329257; PubMed=4578954;  
RA Bridgen J., Kolb E., Harris J.I.;  
RT "Amino acid sequence homology in alcohol dehydrogenase";  
RL FEBS Lett. 33:1-3(1973).  
RN [3]  
RP SEQUENCE OF 34-54.  
RX MEDLINE=79169263; PubMed=436831;  
RA Jeck R., Woenckhaus C., Harris J.I., Runswick M.J.;  
RT "Identification of the amino acid residue modified in Bacillus  
stearothermophilus alcohol dehydrogenase by the NAD+ analogue 4-(3-  
bromoacetylpyridinio)butyldiphosphoadenosine";  
RL Eur. J. Biochem. 93:57-64(1979).  
RN [4]  
RP SEQUENCE OF 1-37; 188-197; 247-263 AND 324-336.  
RC STRAIN=NCA 1503;  
RX MEDLINE=94325354; PubMed=8049268;  
RA Robinson G.A., Bailey C.J., Dowds B.C.A.;  
RT "Gene structure and amino acid sequences of alcohol dehydrogenases of  
Bacillus stearothermophilus";  
RL Biochim. Biophys. Acta 1218:432-434(1994).  
CC -1- FUNCTION: THERMOSTABLE NAD(+) DEPENDENT ALCOHOL DEHYDROGENASE.  
CC -1- CATALYTIC ACTIVITY: An alcohol + NAD(+) = an aldehyde or ketone +  
NADH.  
CC -1- COFACTOR: Binds 2 zinc ions per subunit (By similarity).  
CC -1- ENZYME REGULATION: SUBSTRATE INHIBITION IS NOT OBSERVED WITH ANY  
ALCOHOLS. AND THE ENZYME-NADH DISSOCIATION IS NOT CONSIDERED TO BE  
A RATE-LIMITING STEP.  
CC -1- SIMILARITY: Belongs to the zinc-containing alcohol dehydrogenase  
family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.ebi.ac.uk/announcements/>  
CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
CC -----  
DR EMBL: D90421; BAA14411.1; -;  
DR PIR: A42654; A42654.  
DR InterPro: IPR002328; ADH\_zinc.  
DR InterPro: IPR002085; Adh\_zn\_family.  
DR Pfam: PF00107; ADH\_zinc\_N; 1.  
DR PROSITE: PS00059; ADH\_ZINC; 1.  
KW Oxidoreductase; Zinc; Metal-binding; NAD.  
FT METAL 38 38 ZINC 1 (CATALYTIC) (BY SIMILARITY).  
FT METAL 61 61 ZINC 1 (CATALYTIC) (BY SIMILARITY).  
FT METAL 92 92 ZINC 2 (BY SIMILARITY).  
FT METAL 95 95 ZINC 2 (BY SIMILARITY).  
FT METAL 98 98 ZINC 2 (BY SIMILARITY).  
FT METAL 106 106 ZINC 2 (BY SIMILARITY).  
FT METAL 148 148 ZINC 1 (CATALYTIC) (BY SIMILARITY).  
FT MUTAGEN 40 40 T->S: LITTLE DECREASE IN ACTIVITY.  
FT MUTAGEN 43 43 H->R: HIGHER LEVEL OF ACTIVITY AT PH 9.  
FT MUTAGEN 38 38 C->S: NO ACTIVITY.  
FT MUTAGEN 40 40 T->A: NO ACTIVITY.  
FT MUTAGEN 43 43 H->A: NO ACTIVITY.  
FT CONFLICT 22 22 MISSING (IN REF. 2).  
FT CONFLICT 33 33 MISSING (IN REF. 2).  
FT CONFLICT 52 53 KP -> PK (IN REF. 3).  
SQ SEQUENCE 337 AA; 36100 MW; B9B35A80E9B7A86 CRC64;  
Query Match 7.4%; Score 141.5; DB 1; Length 337;  
Best Local Similarity 24.5%; Pred. No. 0.0047;  
Matches 90; Conservative 57; Mismatches 152; Indels 69; Gaps 18;  
QY 8 KAVIIEDGKAVKT-DVSVPELKEGTALVKVEAVAGNPTDW--KHAYKIGPEGSIL-GC 63  
Db 2 KAAVVEQFKPLQVKEVEKPKISYGEVLVRIKACGVCHTDLHAAHGDMVPKPLPLPGH 61  
QY 64 DIAGTVVKLQPNASTDLKVG-DTGFVHGA-----SQTDPKNGAF 103  
Db 62 EGVGVIEEVGPGV--THLKVDVRGIPWLYSACGCHDYCLSGQETLCERQGNAGVSDGGY 120  
QY 104 AEYAVYVPLFYKSNLTHSTADISEGPVK-NPESASLPVSLTTAGVSLCHHL---GSK 159  
Db 121 AEYCR-----AAADYVVKIPDNLSFEAA-----PIFCAGVTYKALKVTGAK 163  
QY 160 M-EWHPSTPQHTPLLIWGTATVAGQOLQVAKHINAYTKIVTVASKHEKLLKSYGADD 218  
Db 164 PGW-----VAIY-GIGLGHVAVQYAKAMGLNVAVDLGDEKLE-LAKQLGADL 211  
QY 219 VPDYHDAGVTEIQKSKYPNLQHVDAVGSSESIPEAYKVTADSLPATLLEWPMWTIESIP 278  
Db 212 VVNPKHDDAAQWIKKEKVGGVHATVAVSKAAPESAYKIRRGACVVLGLPP---BEIP 268  
QY 279 BEIRKONKVIDITLLYRASG-QEILLGATFPASPVEYHEATVKVFINPHLNGDIHMH 337  
Db 269 IPI-FDTVLNGVKIIGSIVGTRKDLQALQFAA-----EGKVKTIVEVPLENINDVDR 322  
QY 338 NIKVPSNG 345  
Db 323 MLKQING 330  
RESULT 24  
TDH\_PYRAB  
ID TDH\_PYRAB STANDARD; PRT; 348 AA.  
AC Q9UYX0;  
DT 15-SEP-2003 (Rel. 42, Created)  
DT 15-SEP-2003 (Rel. 42, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)



```
FT METAL          99          ZINC 2 (BY SIMILARITY).
FT METAL        107          ZINC 2 (BY SIMILARITY).
FT METAL        148          ZINC 1 (CATALYTIC) (BY SIMILARITY).
FT CONFLICT      54          T -> N (IN REF. 2).
SQ SEQUENCE    341 AA; 3720 MW; AAB2877DD5FAC1E CRC64;

Query Match          7.3%; Score 141; DB 1; Length 341;
Best Local Similarity 20.4%; Pred. No. 0.0052;
Matches 81; Conservative 52; Mismatches 147; Indels 118; Gaps 18;

QY 12 IEQDKAVVKTDSVPPELKEGTALVKVEAVAGNPTD-----WKHIAYKIGPEGSILGCDIA 66
DB 7 LKAEIGIWMTDVPPELGHNDLLIKIRKTAICGTDVHIYNWDEWSQKTIPTVPMVVGHEYV 66
QY 67 GTVVKLGPNASTDLKVGDTGFGFVH-----GASQTDPRNGAF 105
DB 67 GEVVIGIQEVK-GPKIGDRVSGEGHITCGHRCRCGRTHLCRNTIGVGNRP--GCFAE 123
QY 106 YARVYPLFYKSNLTHSTADEISE--GPVKNFESAASLPVSLTTAGVSLCHLGSKMEWH 163
DB 124 YLVI--PAPNAFKIPDNISDLSIFDP---FGNAVHTLSFDLVGED----- 166
QY 164 PSTPQHTPLLIWGATAVGQQLIQVAKHINAYTKIVTASKEKHLKLSYGADDDVFDY- 222
DB 167 -----VLVSGAGPIGIMAAVAKHVGARNVIT-----DVNEYR 200
QY 223 -----HDAGVIEQIKSKYPNLQHVDAVGSBDSI-----PEAYKVTADSL-----PAT 265
DB 201 LELARKNGITRAVNVAKENLNDVMAELGTEGPDVGLMSGAPPAPFTMLDTMNHGRIA 260
QY 266 LLEVPMPTIESIPBEIRKONVKDIT--LLYRASQGEILLGATRPASPPEYHEATVKPVK 324
DB 261 MLGIPP-----SDMSIDWTKVIFKGLFIKGIYGREMF-----ETWYKMAAL 301
QY 325 INPHLNGDI--HHMNKVFNSGLDDVPALTEGKEGK 360
DB 302 IQSGLDLSPIITHRFSIDDFQKGFDP---AMCSG-QSGK 335

RESULT 26
TDH_SHIFL
ID TDH_SHIFL STANDARD; PRT; 341 AA.
AC PS9409;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE L-threonine 3-dehydrogenase [EC 1.1.1.103].
GN TDH OR SF3656.
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=623;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=301 / Serotype 2a;
RX MEDLINE=2272406; PubMed=12384590;
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
RA Yu J.;
RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
RL through comparison with genomes of Escherichia coli K12 and O157."
RL Nucleic Acids Res. 30:4432-4441(2002).
CC -1- CATALYTIC ACTIVITY: L-threonine + NAD(+) = L-2-amino-3-
CC oxobutanoate + NADH.
CC -1- COFACTOR: Binds 2 zinc ions per subunit (By similarity).
CC -1- PATHWAY: Threonine catabolism; first step.
CC -1- SUBUNIT: Homotetramer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: Belongs to the zinc-containing alcohol dehydrogenase
CC family.
-----
```

```
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.1eb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AE015373; AAN45103.1; -.
DR HAMAP; MF 00627; -.
DR PROSITE; PS00059; ADH_ZINC; 1.
KW Oxidoreductase; NAD; Zinc; Metal-binding.
FT METAL 38 38 ZINC 1 (CATALYTIC) (BY SIMILARITY).
FT METAL 63 63 ZINC 1 (CATALYTIC) (BY SIMILARITY).
FT METAL 93 93 ZINC 2 (BY SIMILARITY).
FT METAL 96 96 ZINC 2 (BY SIMILARITY).
FT METAL 99 99 ZINC 2 (BY SIMILARITY).
FT METAL 107 107 ZINC 2 (BY SIMILARITY).
FT METAL 148 148 ZINC 1 (CATALYTIC) (BY SIMILARITY).
SQ SEQUENCE 341 AA; 37255 MW; AAB2877DC84AAC1E CRC64;

Query Match          7.3%; Score 141; DB 1; Length 341;
Best Local Similarity 20.4%; Pred. No. 0.0052;
Matches 81; Conservative 52; Mismatches 147; Indels 118; Gaps 18;

QY 12 IEQDKAVVKTDSVPPELKEGTALVKVEAVAGNPTD-----WKHIAYKIGPEGSILGCDIA 66
DB 7 LKAEIGIWMTDVPPELGHNDLLIKIRKTAICGTDVHIYNWDEWSQKTIPTVPMVVGHEYV 66
QY 67 GTVVKLGPNASTDLKVGDTGFGFVH-----GASQTDPRNGAF 105
DB 67 GEVVIGIQEVK-GPKIGDRVSGEGHITCGHRCRCGRTHLCRNTIGVGNRP--GCFAE 123
QY 106 YARVYPLFYKSNLTHSTADEISE--GPVKNFESAASLPVSLTTAGVSLCHLGSKMEWH 163
DB 124 YLVI--PAPNAFKIPDNISDLSIFDP---FGNAVHTLSFDLVGED----- 166
QY 164 PSTPQHTPLLIWGATAVGQQLIQVAKHINAYTKIVTASKEKHLKLSYGADDDVFDY- 222
DB 167 -----VLVSGAGPIGIMAAVAKHVGARNVIT-----DVNEYR 200
QY 223 -----HDAGVIEQIKSKYPNLQHVDAVGSBDSI-----PEAYKVTADSL-----PAT 265
DB 201 LELARKNGITRAVNVAKENLNDVMAELGTEGPDVGLMSGAPPAPFTMLDTMNHGRIA 260
QY 266 LLEVPMPTIESIPBEIRKONVKDIT--LLYRASQGEILLGATRPASPPEYHEATVKPVK 324
DB 261 MLGIPP-----SDMSIDWTKVIFKGLFIKGIYGREMF-----ETWYKMAAL 301
QY 325 INPHLNGDI--HHMNKVFNSGLDDVPALTEGKEGK 360
DB 302 IQSGLDLSPIITHRFSIDDFQKGFDP---AMRSG-QSGK 335

RESULT 27
ADH2_KLUMA
ID ADH2_KLUMA STANDARD; PRT; 347 AA.
AC QSP4C2;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Alcohol dehydrogenase II (EC 1.1.1.1).
GN ADH2.
OS Kluyveromyces marxianus (Yeast) (Kluyveromyces fragilis).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
OX NCBI_TaxID=4911;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 12424;
RX MEDLINE=20432119; PubMed=10974568;
RA Ladriere J.-M., Georis I., Guerin M., Vandenhaute J.;
RT "Kluyveromyces marxianus exhibits an ancestral Saccharomyces
-----
```

```
RT cerevisiae genome organization downstream of ADH2. ";
RL Gene 255:83-91(2000).
CC -1- CATALYTIC ACTIVITY: An alcohol + NAD(+) = an aldehyde or ketone +
CC NADH.
CC -1- COFACTOR: Binds 2 zinc ions per subunit (By similarity).
CC -1- SUBUNIT: Homotrimer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to the zinc-containing alcohol dehydrogenase
CC family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF225206; AAF91235.1; -.
DR InterPro; IPR002328; ADH_zinc.
DR InterPro; IPR002085; Adh_zn family.
DR Pfam; PF00107; ADH_zinc_N; 1.
DR PROSITE; PS00059; ADH_ZINC; 1.
DR Oxidoreductase; Zinc; Metal-binding; NAD; Multigene family;
KW Acetylation.
FT INIT_MET 0 0 BY SIMILARITY.
FT MOD_RES 43 1 ACETYLATION (BY SIMILARITY).
FT METAL 43 43 ZINC 1 (CATALYTIC) (BY SIMILARITY).
FT METAL 66 66 ZINC 2 (CATALYTIC) (BY SIMILARITY).
FT METAL 97 97 ZINC 2 (BY SIMILARITY).
FT METAL 100 100 ZINC 2 (BY SIMILARITY).
FT METAL 103 103 ZINC 2 (BY SIMILARITY).
FT METAL 111 111 ZINC 2 (BY SIMILARITY).
FT METAL 153 153 ZINC 1 (CATALYTIC) (BY SIMILARITY).
SQ SEQUENCE 347 AA; 36837 MW; F70DD4034FCBC49 CRC64;
Query Match 7.3%; Score 140.5; DB 1; Length 347;
Best Local Similarity 21.5%; Pred. No. 0.0058;
Matches 89; Conservative 54; Mismatches 138; Indels 133; Gaps 20;
QY 2 SVPTTQKAVII-EGDKAVVKTDSVPELKSGTALVKVEAVAGNPTD---WK-----HI 50
Db :||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
1 SIPTTQKGVIFYENGQQLYKDIIPKPKSNELLINVKYSGVCHTDLHAWKGDWPLDTKL 60
QY 51 AYKIGPGSLGCDIAGTVVVLGNASTDLKVED-TGFGFVHGASQTDPK-----99
Db :||| : : : : : : : : : : : : : : : : : : : : : : : : : :
61 PLVGGHEG-----AGVWVAMGDNVK-GWKIGLAGIKWLNGSCMNCCELSNESNCP 112
QY 100 -----NGAPAEYARVYPPPLFYKSNLTHSTADEISEGPVKNFESAASLPVS-----144
Db :||| : : : : : : : : : : : : : : : : : : : : : : : : : :
113 DADLSGYTHDGSFOQYA-----TDAV-----QAHIIPAGTDLAQV 148
QY 145 --LTTAGVSLCHLHLSKM-----EWHPSTPQTHPELLIWWGATAVGQQLIQVAKHINAYTK 198
Db :||| : : : : : : : : : : : : : : : : : : : : : : : : : :
149 APILCAGTVVYKALKTAEMKAGDW-----VAISGAAGGLGSLAVQYAKAM-GFRV 197
QY 199 IVTVASKKHEKLKSYGADVDVHDAGVIEQIKSKYPNLQHVLDVAGSDDSPYAVKVT 258
Db :||| : : : : : : : : : : : : : : : : : : : : : : : : : :
198 LGIDGGEGKEELFKSLGG-EVF-----IDFTKSKDIIVGEVIKAT 235
QY 259 ADSLPAFL-LEVVPMTTETIEPIEIRKDNVKIDITLLYRASGOETLLGATR-PPASPEYHE 316
Db :||| : : : : : : : : : : : : : : : : : : : : : : : : : :
236 NGGAHVGVINSVSEKALESSIE-----YCRSNGTVVLVLGPKDAKCKSDVFN 282
QY 317 ATVKFVKFIPHL-NGGDIHH-----MNKVFNSGDDVDPALTEGIEKEGK 360
Db :||| : : : : : : : : : : : : : : : : : : : : : : : : : :
283 QVVKSIHIVGSYGVNRAETREALDFCKGLGVNAPIKVVGSLTLPPIEYKEVMEQKG 336
RESULT 28
TDH_ECOLI
ID -TDH_ECOLI
AC P07513;
```

```
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE L-threonine 3-dehydrogenase (EC 1.1.1.103).
GN TDH OR B3616.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-20 AND 149-153.
RC STRAIN=K12;
RX MEDLINE=89174812; PubMed=2647748;
RA Aronson B.D., Somerville R.L., Epperly B.R., Dekker E.E.;
RT "The primary structure of Escherichia coli L-threonine
RT dehydrogenase.";
RL J. Biol. Chem. 264:5226-5232 (1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=94316500; PubMed=8041620;
RA Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;
RT "Analysis of the Escherichia coli genome. V. DNA sequence of the
RT region from 76.0 to 81.5 minutes.";
RL Nucleic Acids Res. 22:2576-2586 (1994).
RN [3]
RP ACTIVE SITE, REQUIREMENT FOR ZINC, AND PARTIAL SEQUENCE.
RX MEDLINE=91177850; PubMed=2007567;
RA Epperly B.R., Dekker E.E.;
RT "L-threonine dehydrogenase from Escherichia coli. Identification of
RT an active site cysteine residue and metal ion studies.";
RL J. Biol. Chem. 266:6086-6092 (1991).
CC -1- CATALYTIC ACTIVITY: L-threonine + NAD(+) = L-2-amino-3-
CC oxobutanoate + NADH.
CC -1- COFACTOR: Binds 2 zinc ions per subunit. Cobalt, cadmium and
CC manganese can exchange for zinc.
CC -1- ENZYME REGULATION: Activated by manganese or cobalt ions.
CC -1- PATHWAY: Threonine catabolism; first step.
CC -1- SUBUNIT: Homotrimer.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to the zinc-containing alcohol dehydrogenase
CC family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X06690; CAA29884.1; -.
DR EMBL; U00039; AAB18593.1; -.
DR EMBL; AE000439; AAC76640.1; -.
DR PIR; A33276; DEECTH.
DR ECO2DBASE; G038.8; 6TH EDITION.
DR EcoGene; EGI0993; tdh.
DR HAMAP; MF_00627; -; 1.
DR InterPro; IPR002328; ADH_zinc.
DR InterPro; IPR002085; Adh_zn family.
DR InterPro; IPR000205; NAD_binding.
DR InterPro; IPR004627; Tdh.
DR Pfam; PF00107; ADH_zinc_N; 1.
DR TIGRFAMs; TIGR00692; tdh; 1.
DR PROSITE; PS00059; ADH_ZINC; 1.
DR Oxidoreductase; Zinc; Metal-binding; Manganese; Cobalt; NAD;
DR Complete proteome.
FT METAL 38 38 ZINC 1 (CATALYTIC).
FT METAL 63 63 ZINC 1 (CATALYTIC) (BY SIMILARITY).
FT METAL 93 93 ZINC 2 (BY SIMILARITY).
FT METAL 96 96 ZINC 2 (BY SIMILARITY).
FT METAL 99 99 ZINC 2 (BY SIMILARITY).
FT METAL 107 107 ZINC 2 (BY SIMILARITY).
```



```
FT METAL 148 148 ZINC 1 (CATALYTIC) (BY SIMILARITY).
SQ SEQUENCE 341 AA; 37239 MW; 039F6D6B1CE8C2B2 CRC64;

Query Match
Best Local Similarity 7.3%; Score 140; DB 1; Length 341;
Matches . 80; Conservative 52; Mismatches 150; Indels 114; Gaps 17;

Qy 12 IEGDKAVKVTDSVPELKEGTALVKVEAVAGNPTD-----WKHAYKIGPEGSILGCDIA 66
Db 7 LKAEIGWMTDVPPELGHNDLLIKIRKTAICGTDVHIYNWDEWSQKTIIPVPMVVGHEYV 66

Qy 67 GTTVKLGPNASTDLKVGDTGFGFVH-----GASQTDPPKNGAPAE 105
Db 67 GEVVDIGQEVK-GFKIGDRVSGEGHITCGHCRNCRGRTHLCRNTIGVGNRP--GCPAE 123

Qy 106 YARVPPPLFYKSNLTHSTADRISEGPVKNPESAASLPVSLTAGVSLCHLHGSKMEWHP 165
Db 124 YLVI--PAFNAFKPDNISDILA-AIFDPFGNAVHTALSPDLVGED----- 166

Qy 166 TPQHTPLLLINGGATAVGQQLIQVAKHINAYTKIVTASKKHEKLLKSYGADDDVFDY--- 222
Db 167 -----VLVSGAGPIGIMAAVAKHVGARNVIT-----DVNEYRLE 202

Qy 223 --HDAGVTEQIKSKYPNLQHVDAVGSSESI-----PEAVKVTADSL-----PATLL 267
Db 203 LARKMGITRAVNAVAKENLNDVMAELMGTEGFDVGLMSGAPPAPFRMLDTMNHGGRIAML 262

Qy 268 EWPMTIESIFEEIRKONKVKIDIT--LLYRASSQELLGATRPASPSPHEATVFKVFKN 326
Db 263 GIPP-----SMDSDTWKVIKFGIFKIGIYREMP-----ETWYKMAALIQ 303

Qy 327 PHLNNGDI--HHMNIKVFNSGLDDVPALTEGIEKKGK 360
Db 304 SGLDLSPIITHRESIDDFQKGF-----AMRSG-QSGK 335

RESULT 29
TDH_SHEON
ID TDH_SHEON STANDARD; PRT; 341 AA.
AC Q8E8J1,
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE L-threonine 3-dehydrogenase (EC 1.1.1.103).
GN TDH OR S04673.
OS Shewanella oneidensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadaceae; Shewanella.
OX NCBI_TaxID=70863;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=MR-1,
RX MEDLINE=22297686; PubMed=12368813;
RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
RA Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,
RA Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
RA Vanathavan J., Weidman J., Imprial M., Lee K., Berry K., Lee C.,
RA Mueller J., Khouri H., Gill J., Utterback T.R., McDonald L.A.,
RA Feilblyum T.V., Smith H.O., Venter J.C., Neelson K.H., Fraser C.M.,
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
RT Shewanella oneidensis."
RL Nat. Biotechnol. 20:1118-1123(2002).
CC -I- CATALYTIC ACTIVITY: L-Chreonine + NAD(+) = L-2-amino-3-
CC oxobutanate + NADH.
CC -I- COFACTOR: Binds 2 zinc ions per subunit (By similarity).
CC -I- PATHWAY: Threonine catabolism; first step.
CC -I- SUBUNIT: Homotetramer (By similarity).
CC -I- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -I- SIMILARITY: Belongs to the zinc-containing alcohol dehydrogenase
CC family.
CC -----
```

```
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.ebi.ac.uk/announcements/
or send an email to license@ebi.ac.uk).

EMBL; AE015899; AAN57632.1; -.
DR TIGR; SO4673; -.
DR HAMAP; MF_00627; -.
DR InterPro; IPR002328; ADH_zinc.
DR InterPro; IPR002085; Adh zn family.
DR InterPro; IPR002025; NAD_binding.
DR InterPro; IPR004627; Tdh.
DR Pfam; PF00107; ADH_zinc_N; 1.
DR TIGRFAMs; TIGR00692; tdn; 1.
DR PROSITE; PS00059; ADH_ZINC; 1.
KW Oxidoreductase; NAD; Zinc; Metal-binding; Complete proteome.
FT METAL 38 38 ZINC 1 (CATALYTIC) (BY SIMILARITY).
FT METAL 63 63 ZINC 1 (CATALYTIC) (BY SIMILARITY).
FT METAL 93 93 ZINC 2 (BY SIMILARITY).
FT METAL 96 96 ZINC 2 (BY SIMILARITY).
FT METAL 99 99 ZINC 2 (BY SIMILARITY).
FT METAL 107 107 ZINC 2 (BY SIMILARITY).
FT METAL 148 148 ZINC 1 (CATALYTIC) (BY SIMILARITY).
SQ SEQUENCE 341 AA; 37255 MW; 5378A45011537BF CRC64;

Query Match 7.3%; Score 140; DB 1; Length 341;
Best Local Similarity 21.4%; Pred. No. 0.0061;
Matches 81; Conservative 59; Mismatches 138; Indels 100; Gaps 20;

Qy 12 IEGDKAVKVTDSVPELKEGTALVKVEAVAGNPTD-----WKHAYKIGPEGSILGCDIA 66
Db 7 LKAEIGWMTDVPPELGHNDLLIKIRKTAICGTDVHIYNWDEWSQKTIIPVPMVVGHEYV 66

Qy 67 GTTVKLGPNASTDLKVGDTGFGFV-----HGASQTD-----PKNGAPAEYA 107
Db 67 GEVVDIGQEVK-RGFNIGDRVSGEGHITCGHCRNCRGRTHLCRNTSGVGNREGSPAEL 125

Qy 108 RVTPLPYKSNLTHSTADEISE--GPVKN--FESASL-----PVSLLTAG-----VSLC 153
Db 126 VI--PAFNAFKIPDDISDILASIPDFGNVHTALSFDLVGEDVLLITGAGPIGIMAAV 183

Qy 154 HHLGSKMEWHPSTPQHTPLLLINGGATAVGQQLIQVAKHINAYTKIVTASKKHEKLLK 213
Db 184 RHVGAR-----HVV-----TDVNEYRELARKMA-TRAVNAVQENLKDVNKE 226

Qy 214 YGADDDVFDYHDAGVIEQIKSKYPNLQHVDAVGSSEDS-IPEAYKVTADSLPATLLEVP 272
Db 227 LGWTEGFD-----VGLMSGVPSAPRAMLD-----TMNHGKI 259

Qy 273 TISIP-EETRKONVKIDITLLYRASSQELLGATRPASPSPHEATVFKVPKPNPHLN 331
Db 260 AMLGIPGEMADWSKY-----IFKGLVIKIGIYREMP-----ETWYKMAASLQSGLDI 308

Qy 332 GDI--HHMNIKVFNSGLD 347
Db 309 SPIITHYKIDDFQKGF 326

RESULT 30
VATL_TORCA
ID VATL_TORCA STANDARD; PRT; 379 AA.
AC P19333;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Synaptic vesicle membrane protein VAT-1.
OS Torpedo californica (Pacific electric ray).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC Elasmobranchii; Squala; Hypnosquala; Pristiogalea; Batoida;
OC Torpediniformes; Torpedinoidei; Torpedinidae; Torpedo.
```

Search completed: January 10, 2004, 10:05:32  
Job time : 29 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 10, 2004, 09:49:00 ; Search time 84 Seconds  
(without alignments)

1136.660 Million cell updates/sec

Title: US-10-081-644-2

Perfect score: 1920

Sequence: 1 MSVPTTKAVIIIEGKAVK.....ALTEGIKEGKNVYVARL 370

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 100%

Listing first 90 summaries

Database :

SPTREMBL 23:\*\*

- 1: sp\_archaea:\*\*
- 2: sp\_bacteria:\*\*
- 3: sp\_fungi:\*\*
- 4: sp\_human:\*\*
- 5: sp\_invertebrate:\*\*
- 6: sp\_mammal:\*\*
- 7: sp\_mhc:\*\*
- 8: sp\_organelle:\*\*
- 9: sp\_phase:\*\*
- 10: sp\_plant:\*\*
- 11: sp\_rodent:\*\*
- 12: sp\_virus:\*\*
- 13: sp\_vertebrate:\*\*
- 14: sp\_unclassified:\*\*
- 15: sp\_rvirus:\*\*
- 16: sp\_bacteriap:\*\*
- 17: sp\_archaeap:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	309.5	16.1	348	3 Q9P7F4	Q9P7F4 schizosacch
2	294	15.3	362	3 Q9J0F9	Q9J0F9 penicillium
3	277.5	14.5	339	10 Q941I0	Q941I0 fragaria an
4	271	14.1	337	10 Q23939	Q23939 fragaria ve
5	266.5	13.9	363	3 Q9Y7D0	Q9Y7D0 aspergillus
6	264	13.8	313	16 Q8Y9B9	Q8Y9B9 listeria mo
7	261	13.6	386	10 Q943P3	Q943P3 arabidopsis
8	260	13.5	313	16 Q92E39	Q92E39 listeria in
9	259	13.5	317	10 Q43677	Q43677 vigna radia
10	248.5	12.9	309	10 Q8H2V0	Q8H2V0 oryza sativ
11	248	12.9	309	10 Q94F67	Q94F67 helianthus
12	247	12.9	309	10 Q8LEB8	Q8LEB8 arabidopsis
13	219	11.4	311	16 Q8ES25	Q8ES25 oceanobacil
14	218	11.4	353	16 Q8UK00	Q8UK00 agrobacteri
15	214	11.1	313	16 Q85702	Q85702 streptomyce
16	208	10.8	324	16 Q9KFW8	Q9KFW8 bacillus ha

Q8tam3 methanosaer  
Q8h0m1 epinacia ol  
Q67374 aquifex aeo  
Q8dv65 streptococ  
Q98h13 rhizobium l  
Q8f1n0 corynebacte  
Q912a5 streptomyce  
Q9k6b8 bacillus ha  
Q8u1m9 agrobacteri  
Q8ayv1 triphyearia  
Q8cu46 staphylococ  
Q9r1r7 streptomyce  
Q97v32 sulfolobus  
Q8ukd5 agrobacteri  
Q9chl2 lactococcus  
Q9kew1 bacillus ha  
Q8pc20 xanthomonas  
Q92r74 rhizobium m  
Q8ugn9 agrobacteri  
Q8fyw9 brucella su  
Q8y1v7 brucella me  
Q8j284 podospora a  
Q8pnp3 xanthomonas  
Q98j41 rhizobium l  
Q8nid8 corynebacte  
Q98xf6 lithospermu  
Q8lba4 arabidopsis  
Q9l96 arabidopsis  
Q96y05 sulfolobus  
Q98v68 arabidopsis  
Q8uh60 agrobacteri  
Q9kug9 vibrio chol  
Q8t122 dictyosteli  
Q53146 mycobacteri  
Q8y2p5 anabaena sp  
Q9bra4 homo sapien  
Q985b6 rhizobium l  
Q98f10 rhizobium l  
Q8ejr1 shewanella  
Q45496 caenorhabdi  
Q8ruh4 oryza sativ  
Q8pp94 xanthomonas  
Q54297 streptomyce  
Q54296 streptomyce  
P72043 mycobacteri  
Q8p362 xanthomonas  
Q8dc41 vibrio vuln  
Q92851 rhizobium m  
Q9xa55 streptomyce  
Q42703 candida par  
Q8dmw0 synchococc  
Q8zfp1 yersinia pe  
Q8d0p7 yersinia pe  
Q74822 schizosacch  
Q65423 arabidopsis  
Q8pek0 xanthomonas  
Q8wv3 homo sapien  
Q8zup0 pyrobaculum  
Q81616 cryptocospori  
Q91679 homo sapien  
Q9bwb8 homo sapien  
Q9rv98 deinococcus  
Q8ntu9 corynebacte  
Q8crj7 staphylococ  
Q91x24 arabidopsis  
Q81a26 arabidopsis  
Q914j8 pseudomonas  
Q25222 leishmania  
Q8ld14 arabidopsis  
Q92bk1 streptomyce  
Q97wa1 sulfolobus  
Q9c677 arabidopsis  
Q96r18 homo sapien

90 146.5 7.6 484 13 Q8JFV8 Q8Jfv8 brachydanio

## ALIGNMENTS

## RESULT 1

Q9P7F4 PRELIMINARY; PRT; 348 AA.  
AC Q9P7F4  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Putative dehydrogenase by similarity.  
GN SPAC2E1P3.01.  
OS Schizosaccharomyces pombe (Fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomycetes.  
OX NCBI\_TaxID=4896;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=972h-;  
RA Harris D., Wood V., Rajandream M.A., Barrell B.G.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL162531; CAB83005.1; -;  
DR GeneDB SPombe: SPAC2E1P3.01; -;  
DR InterPro; IPR002085; Adh\_zn\_family.  
DR Pfam; PF00107; adh\_zinc; 1.  
SQ SEQUENCE 348 AA; 37929 MW; 63E638A712AFF1C2 CRC64;

Query Match 16.1%; Score 309.5; DB 3; Length 348;  
Best Local Similarity 28.2%; Pred. No. 1.7e-15;  
Matches 110; Conservative 49; Mismatches 160; Indels 71; Gaps 13;

Qy 8 KAVIEGDKAV-VKTDVSVPELKEGTALVKVEAVAGNPTDVKHAYKIGPEGSILGCDIA 66  
Db 2 KAVIADGQNGVEVISDAPKPTPEKGEFLGRVIRVAFNPIDWKTLNLYASIEKGTGGTDFV 61

Qy 67 GTVVKLGPNASTDLKVGDTGTFGVHGSQTDPKNGAFAYARVVPPLFY--KSNLTHSTA 124  
Db 62 AVVEDVGEGVDRSYKIGATVSGWAPG--PLDGSNAAWREYITLDVNLVYFVPKNITPS-- 117

Qy 125 DEISEGPKNFESAASLPVSLTTAGVSLCHHLG-----SKMEWHPSTPQHTHP 172  
Db 118 -----QAATPLPTFTASQGLNQYLGLPPLPTDGSKNASQOKW----- 155

Qy 173 LLIWGGATAVGQQLIQVAKHINAYTKIVTASKEKELLSYGADDVFDYHDAGVIRQIK 232  
Db 156 VLWWSGSSVQYVVLAAH--AGYKVIATCSPHNFDWIKLGADFTVDYDPPNVVEIK 213

Qy 233 SKYNLQHVLDVAGSERSIPAYKVTADSLPATLLEVP-----MTIESIPPEI 281  
Db 214 K-----ATDDSV--FYGDAASFETSTLAVKAFSSKVDGKLINILSSPPSP 259

Qy 282 RKNVNDITLLYRASGOEILLCATRPASPEVHEATVKFVPINPHANGDIIHNNIK 340  
Db 260 RSE-VKLIIGIDYSLFRNFFNFKIEPTQASTDHA-VEYKILGTWLGQEVIIPNRVK 317

Qy 341 VFSNGLDDVPALTGEGKEGKNKVKYVARL 370  
Db 318 EFDGGLQAIKALREFASGKHSVAKFVVRI 347

## RESULT 2

Q8J0F9 PRELIMINARY; PRT; 362 AA.  
AC Q8J0F9  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Enoyl reductase.  
GN MLCG.

OS Penicillium citrinum.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
OC Eurotiiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.  
OX NCBI\_TaxID=5077;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22162865; PubMed=12172803;  
RA Abe Y., Suzuki T., Ono C., Iwamoto K., Hosobuchi M., Yoshikawa H.;  
RT "Molecular cloning and characterization of an ML-236B (compactin)  
biosynthetic gene cluster in Penicillium citrinum.";  
RL Mol. Genet. Genomics 267:636-646(2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=2227327; PubMed=12242508;  
RA Abe Y., Suzuki T., Mizuno T., Ono C., Iwamoto K., Hosobuchi M.,  
Yoshikawa H.;  
RT "Effect of increased dosage of the ML-236B (compactin) biosynthetic  
gene cluster on ML-236B production in Penicillium citrinum.";  
RL Mol. Genet. Genomics 268:130-137(2002).  
DR EMBL; AB072893; BAC20562.1; -;  
SQ SEQUENCE 362 AA; 39600 MW; CDB2CC5EFBED0E0F CRC64;

Query Match 15.3%; Score 294; DB 3; Length 362;  
Best Local Similarity 26.9%; Pred. No. 2.8e-14;  
Matches 104; Conservative 66; Mismatches 160; Indels 56; Gaps 15;

Qy 3 VPTQKAVIIEGDK--AVVKTDVSVPELKEGTALVKVEAVAGNPTDVKHAYKIGPEGS 60  
Db 11 IPPVKQTILTNDKDEWIMDDAPTCKLPADQYVRIHAVVNPSTK-MRGDFATPFAC 69

Qy 61 LGCDIAGTVVVKLGPNASTDLKVGDTGTFGVHGSQTDPKNGAFAYARVVPPLFYKSNLT 120  
Db 70 LGTDYAGTVVAVG--SEITHVKVGDRVFGAGNEMCPRTPEQAGAFSQT----- 115

Qy 121 HSTADRISEGPV-----KNFESAASLPVSLTTAGVSLCHHLGSKMEWHPSTPQHTPL 173  
Db 116 -----ITRGIWAKIPDSMTWEAASLPAGISTTGLAM-KLGLMPLPYSETKPSKTYV 168

Qy 174 LLIWGGATAVGQQLIQVAKHINAYTKIVTASKEKELLSYGADDVFDYHDAGVIRQIKS 233  
Db 169 LIYGSSTATATIAMQFMR--LSGYTPIAT--CSHKNFDLAKNGAEEVFDYRDADCAQKIRD 226

Qy 234 -KYPNLQHVLDVAGSERSIPAYKVTADSLPATLLEVPMTIESIPPE-IRKDNVKIDIT 291  
Db 227 YTRNNLAYALDCIINVESTCYKATGRAGRY-----VALNPFPEHAATKMTSDWT 280

Qy 292 LLYRASGOEILLCATRPASPEVHEATVKFVPINPHL-----NNGDIHNNIKVFS 343  
Db 281 L-----GPTIFGSGTWPA-PYGCEASEE-VRLFGTQLQVASRLVEEDKLYHHPLRVID 333

Qy 344 NGLDDVPALTGEGKEGKNKVKYVAR 369  
Db 334 GGLEQVQKQGMETVRNGELSGEKIVVR 359

## RESULT 3

Q94110 PRELIMINARY; PRT; 339 AA.  
AC Q94110  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
DE Putative quinone oxidoreductase (Fragment).  
OS Fragaria ananassa (Strawberry).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids I; Rosales; Rosaceae; Rosoideae; Fragaria.  
OX NCBI\_TaxID=3747;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Chandler;  
RA Lopez-Raez J.A., Redondo-Nevado J., Medina-Escobar N.,  
Caballero-Repullo J.L., Munoz-Blanco J.;

QY 53 KIGEGSILGCDIAGTVMKLGPNASTDLKVGDTGFGFVHGASQDTP-KNGFAFAEYRVYP 111

Db 281 -----LGPTIFGEGSTWPAPYKPCGSEBERQFGEDLWRITAGQLVEDGRLVHHPL 329

```
QY 340 KVFSGNGLDDVPALTEGIGKGNKXKVVYARL 370
Db 330 RVQVGFDHKKQGMELVRKGLSGEKLVLRL 360

RESULT 6
Q8Y9B9 PRELIMINARY; PRT; 313 AA.
AC Q8Y9B9, 2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein lmo0613.
GN Lmo0613.
OS Listeria monocytogenes.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxID=1639;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EGD-e / Serovar 1/2a;
RX MEDLINE=21537279; PubMed=1679669;
RA Glaser P., Frangeul L., Buchrieser C., Ruiniok C., Amend A.,
RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussauguet O.,
RA Etian K.-D., Faihi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaeft U., Krefit J., Kuhn M., Kunst F., Kurapat G.,
RA Madueno E., Maitouram A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordsiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schluter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
DR EMBL; ALS91976; CAC98691.1; -
DR Llistalist; LMO0613; -
DR InterPro; IPR002085; Adh_zn_family.
DR Pfam; PF00107; adh_zinc; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 313 AA; 33940 MW; 0F1012C59A23621D CRC64;

Query Match 13.8%; Score 264; DB 16; Length 313;
Best Local Similarity 29.08; Pred. No. 4.4e-12;
Matches 102; Conservative 56; Mismatches 120; Indels 74; Gaps 18;

QY 8 KAVIIE---GDKAVKTDVSVPELKEGTALVKVEAVAGNPTDWK----HIAYKIGPE-GS 59
Db 2 KAVVIENYGKKEELKEKEVAMPKAGNQIVVKEAATSINDWKREGYLKQMWDFPI 61

QY 60 ILGCDIAGTVVVLGPNASTDLKYDGTGFGFVHGASQTDPKNGAFAEYARYVPLFYKSNL 119
Db 62 ILGMDVAGVISEVGEVY-TDWKVGDEYF-----ARPETTFGTAYEATV-----DDHL 109

QY 120 THSTADISEGPKVKNFESAASLPVSLTTAGVSLCHLGSXMEWHPSPQHTHPLLIWGGA 179
Db 110 LAPLEGIS-----FDEAASIPLAGUTAWQALFDH--AKL-----QKGEKVLHAGA 154

QY 180 TAVGQQLIQVAKHINAYTKIVTVASKKHEKLLSYGADDDYFDHAGVIEQIKSKYPNLQ 239
Db 155 GGVGTLAIQLAKAGA--EVITTASAKNHELLSLGADQVIDYKEV---NFKDVLSDID 208

QY 240 HVIDAVGSEDSIEAYKVTADSLPATLLEVVPMTIESIPEIRKDNVDKIDITILLY-RASG 298
Db 209 VVFTMGQGIIE-TDSYDVLKEG-TGRLVSIIV-----GISNEDRAKEKNVTATGILQPNG 261

QY 299 QEI-----LLGATRFP-----ASPEYHEATVVKFVKFN 326
Db 262 EQLKELGKLLANKTIKPIVGTAT-FPPEKGVGFDHALSETTHAVGKIVISFN 312

RESULT 7
Q945P3 PRELIMINARY; PRT; 386 AA.
AC Q945P3, 2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE AT123740/F508.27.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosid1; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Shinn P., Chen H., Cheuk R., Kim C.J., Koesema E., Meyers M.C.,
RA Banh J., Bowser L., Carninci P., Dale J.M., Goldsmith A.D.,
RA Hayashizaki Y., Ishida J., Jiang P.X., Jones T., Kamiya A.,
RA Karlin-Neumann G., Kawai J., Lam B., Lee J.M., Lin J., Liu S.X.,
RA Miranda M., Narusaka M., Nguyen M., Onodera C.S., Palm C.J.,
RA Pham P.K., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,
RA Tang C.C., Toriumi M., Yamada K., Yamamura Y., Yu G., Yu S.,
RA Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RT "Arabidopsis cDNA clones.";
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Cheuk R., Chen H., Kim C.J., Meyers M.C., Shinn P., Banh J.,
RA Bowser L., Carninci P., Chang E., Dale J.M., Goldsmith A.D.,
RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Lam B., Lee J.M., Lin J., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Quach H.L., Sakurai T., Satou M.,
RA Seki M., Southwick A., Tang C.C., Toriumi M., Wu H.C., Yamada K.,
RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
RA Ecker J.R.;
RT "Arabidopsis ORF clones.";
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF411799; AAL06488.1; -
DR EMBL; AY094032; AAM16188.1; -
DR InterPro; IPR002085; Adh_zn_family.
DR Pfam; PF00107; adh_zinc; 1.
SQ SEQUENCE 386 AA; 40986 MW; 4DD19133C9D5D84 CRC64;

Query Match 13.6%; Score 261; DB 10; Length 386;
Best Local Similarity 32.2%; Pred. No. 1e-11;
Matches 88; Conservative 37; Mismatches 104; Indels 44; Gaps 11;

QY 2 SVPTTKQAVIIEG---DXAVKTDVSVPELKEGTALVKVEAVAGNPTDWKHIAYKI--- 54
Db 73 SIPKEMKAWVSYDYGVDVLYKLESNTVWPEIKEDQVLIKVAAALNFPVDAKRQGRFKAT 132

QY 55 -GPEGILGCDIAGTVVVLGPNASTDLKYDGTGFGFVHGASQTDPKN-GAFAEYARYVPP 112
Db 133 DGPLTPVPGDVAGVVKVVG-SAVKDLKEGDEVYANVSEKALEGPKQFGLSLAYTAEEK 191

QY 113 L--FYKSNLTHTSTADEISEGPKVKNFESAASLPVSLTTAGVSLCHLGSXMEWHPSPQHT 170
Db 192 LLALKPKNI-----DFAQAAGLPLAIETADEGLV-----RTEFSAG----- 227

QY 171 HPLIINGGATACVQQLIQVAKHINAYTKIVTVASKKHEKLLSYGADDDYFDHAGVIEQ 230
Db 228 KSLVLNAGGAGVGLVLIQAKHYGASKVAATASTEKLELVRLSLGADLAIDYTKEN-IED 286

QY 231 IKSYPENLQHVDAVGSSEDS---IPEAYKVTA 259
Db 287 LPDKY---DVFDAGMCDKAVKVIKEGGKWA 316

RESULT 8
Q92E39 PRELIMINARY; PRT; 313 AA.
AC Q92E39, 2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
```











"The genome of Methanosarcina acetivorans reveals extensive metabolic RT and physiological diversity.";  
RL Genome Res. 12:532-542(2002).  
DR EMBL; AE010739; AAM04212.1; -  
DR InterPro; IPR002085; Adh\_zn\_family.  
DR InterPro; IPR002364; QOR\_zeta\_crystal.  
DR Pfam; PF00107; adh\_zinc; 1.  
DR PROSITE; PS01162; QOR\_ZETA\_CRYSTAL; 1.  
KW Complete proteome.  
SQ SEQUENCE 309 AA; 33158 MW; D56D002077DA241D CRC64;

Query Match 10.4%; Score 200; DB 17; Length 309;  
Best Local Similarity 26.2%; Pred. No. 3.5e-07;  
Matches 79; Conservative 52; Mismatches 120; Indels 50; Gaps 15;

Qy 8 KAVIIE--GDKAVVKTDSVPPELKEGTALVKEAVAG--NPTDWKHIAKIG--PEGSIL 61  
Db 2 KAIRIEFGGKEVKNYE--DIPOPGTGEIRIRIAGVNPMDWKIRSGIMENPLPMTM 60  
Qy 62 GCDIAGTVVKGPNASTDLKVGDTGFGFVHGASQTDPKNGAFAYARYVPPFLFYKSNLTH 121  
Db 61 GIDVAGVVDAGQP--GEVSFGQGEVVF-----AKVSGQGSYASYTVV----- 101  
Qy 122 STADEISEGPVK--NFESAASLPVSLTTAGVSLCHLGSKMEWHFSTQHTHPLIWGGAT 180  
Db 102 -NSAQVARKPKSIGFVESAALPTAGLAWSQSLFIAGL-----EKQSVLIHGAAG 151  
Qy 181 AVGQQLQVAKHINAYTKIVTVASKKHEKLKSYGADDPVFDHAGVTEQIKSKYVPLQH 240  
Db 152 GVGSAFVQFAWKQAY--VICTASEKNQFLKSLGTDEFIDYKK-----QRFEDAGKVDV 205  
Qy 241 VIDAVGSEDSIPEAYKYVADSLPATLLEVPMTTIESIPEIRKONVKNIDITLLYRASQGE 300  
Db 206 VLDITGG--DTEDRSWGLKPG--GFLVSTVARIPEGVPE---KYGVAHQ--TLMTADGEE 258  
Qy 301 I 301  
Db 259 L 259

RESULT 18  
Q8HOM1 PRELIMINARY; PRT; 329 AA.  
AC Q8HOM1  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Quinone-oxidoreductase homologue.  
GN QOR.  
OS Spinacia oleracea (Spinach).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
OC Caryophyllidae; Caryophyllales; Chenopodiaceae; Spinacia.  
OX NCBI\_TaxID=3562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Leaf;  
RA Mitas S., Salvi D., Ferro M., Grunwald D., Garin J., Joyard J.,  
RA Rolland N.;  
RT "Non canonical transit peptide for import into the chloroplast.";  
RL J. Biol. Chem. 0:0-0(0).  
DR EMBL; AJ511792; CAD54431.1; -.  
SQ SEQUENCE 329 AA; 34811 MW; 908A03516336890E CRC64;

Query Match 10.4%; Score 199.5; DB 10; Length 329;  
Best Local Similarity 29.0%; Pred. No. 4.2e-07;  
Matches 74; Conservative 34; Mismatches 112; Indels 35; Gaps 9;

Qy 14 GDKAVVKTDSVPPELKEGTALVKEAVAGNPTDWKHIAKIGP-----EGSILGCDIAGT 68  
Db 17 GTDALKHVEAVAPDPKSDLLKLEATNLNFDWKIKQGVLRPLLPKFPPTIGTDVAGE 76  
Qy 69 VVKLGNASTDLKVGDTGFGFVHGASQTDPKNGAFAYARYVPPFLFYKSNLTHSTADEIS 128

Query Match 10.1%; Score 193; DB 16; Length 343;  
Best Local Similarity 28.7%; Pred. No. 1.4e-06;  
Matches 77; Conservative 43; Mismatches 94; Indels 54; Gaps 14;

Qy 8 KAVIIEG----DKAVVKTDSVPPELKEGTALVKEAVAGNPTD--WKHI--AYKIGPE--GSI 60  
Db 2 KAVILTGFGGIENLKYVEDFPKPEKDEVLIRKVAVALNHLDIWVRMGALPVKPELPHI 61  
Qy 61 LGCDIAGTVVKGPNASTDLKVG-----DTGFGFVHGASQTD-----PQNG 101  
Db 62 LGSDVSGVKEVG--SLVKNVKEGEVIIAPGLSCGVCWDCQSGRDNHCKDYDILGLKNKG 120  
Qy 102 AFREYARVYPPFLFYKSNLTHSTADEISEGPKVN--PESAASLPVSLTTAGVSLCHLGSK 159  
Db 121 GYAEYVAV-----PARNVIKRP--KNLSFEAAASYPLTFLT-----VNALVDK 162  
Qy 160 MEWHFSTQHTHPLIWGGATAVGQQLIQVAKHINAYTKIVTVASKKHEKLKSYGADDPV 219  
Db 163 AQIKPYS-----RVFIWAGSSGVGAGIQLAKAFNAFV--ITTAGNEEKAKCKELGADLV 216  
Qy 220 FDYHDAGVTEQIKSKY--PNLQHVIDAVG 246  
Db 217 FNHYKDDVVKKREVFPKGEVDVVVDHIG 244

RESULT 20  
Q8DV65 PRELIMINARY; PRT; 322 AA.  
ID Q8DV65



```
Db 109 ---HLVDERLVGHAPGSV-DATHAALPLVFLTAWEALFDRL-----PVTQSTGTGLL 157
Qy 175 IWGATAVGQQLIQVAKHINAYTKIVTVASKKHEKLLKSYGADVDVHDAGVIEQIKSK 234
Db 158 VLGGAGVPSALIQALRALTLNLRVATASRESTRWVREMGANDVIDHH-----HDLREQ 212
Qy 235 YPNLOHVUDA--VGSEDSI-----PEAYKVTDAD 260
Db 213 ISGVDFVFSWTTGREBELTAVMSPQSHLVLD 245

RESULT 23
Q9L2A5
ID Q9L2A5 PRELIMINARY; PRT; 339 AA.
AC Q9L2A5;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Putative zinc-binding oxidoreductase.
GN SC05417 OR SC8F4.21.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycinae; Streptomycetaceae; Streptomycetes.
OX NCBI_TaxID=1902;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=A3(2);
RC STRAIN=A3(2);
RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
SEQUENCE FROM N.A.
RP STRAIN=A3(2);
RC STRAIN=A3(2);
RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
SEQUENCE FROM N.A.
RP STRAIN=A3(2);
RC STRAIN=A3(2);
RA Redenbach M., Kieser H.M., Denapaita D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RA "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
RN [4]
SEQUENCE FROM N.A.
RP STRAIN=A3(2) / M145;
RC STRAIN=A3(2) / M145;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.F., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RL coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL; AL939123; CAB70647.1; -.
DR HSSP; P28304; 1QOR.
DR InterPro; IPR002085; Adh zn family.
DR DRIPRO; IPR002364; OOR_zeta_crystal.
DR Pfam; PF00107; adh_zinc.1.
DR PROSITE; PS01162; OOR_ZETA_CRYSTAL; 1.
KW Complete proteome.
SQ SEQUENCE 339 AA; 34913 MW; A8A43FB244A5C7C8 CRC64;

Query Match 9.8%; Score 188; DB 16; Length 339;
Best Local Similarity 28.5%; Pred. No. 3.4e-06;
Matches 71; Conservative 30; Mismatches 90; Indels 58; Gaps 10;

Qy 14 GDKAVVKTDSVPPELKEGTALVKVEAVAGNPTDWKHIAKYI--GPEGSILGCDIAGTVVK 71
```

```
Db 41 GPEVRLTELPAPGGEILVAVHAAGLNTDFKRALKIFLPPPLTLGWDVSGTVE 100
Qy 72 LGPNASTDLKVGDTGFG---FVHGASQTDPKNGAFABYARVYPLFYKSNLTHSTADEIS 128
Db 101 TGFQV-TLFRPGDEVFGMLPYPHG-----HGHABYV----- 131
Qy 129 EGPVKNFES-----AASLPVSLTTAGVSLCHLGSXMEWHPSTPQHTHPLLWGGA 179
Db 132 TGPAPAFAPAGCTDHHVQAAALPLAALTAWQALVDTAGLRA-----GQVLIHAAA 182
Qy 180 TAVGQQLIQVAKHINAYTKIVTVASKKHEKLLKSYGADVDVHDAGVIEQIKSKYNLQ 239
Db 183 GGVGHVLAQVIAKERGAH--VTGTASAPKHAFLRELGAACVDRSADFTD--TEERY--D 236
Qy 240 HVIDAVGSE 248
Db 237 VLDDTLGGE 245

RESULT 24
Q9KEB8
ID Q9KEB8 PRELIMINARY; PRT; 322 AA.
AC Q9KEB8;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE Quinone oxidoreductase.
GN B09335.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RL halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
DR EMBL; AP001510; BAB04654.1; -.
DR InterPro; IPR002085; Adh zn family.
DR Pfam; PF00107; adh_zinc.1.
KW Complete proteome.
SQ SEQUENCE 322 AA; 34771 MW; B9D16F204BC5CE2E CRC64;

Query Match 9.8%; Score 187.5; DB 16; Length 322;
Best Local Similarity 27.0%; Pred. No. 3.4e-06;
Matches 79; Conservative 44; Mismatches 83; Indels 87; Gaps 14;

Qy 8 KAVIIE--GKAVVKT-DVSVPELKEGTALVKVEAVAGNPTD-----WKHIAY 52
Db 2 KAVVYEQGDSVVLKVMVDPKPTIGPTDVLINVKASGTPNVDYFRKGIRQVEQPHIP 61
Qy 53 KIGPEGSILGCDIAGTVVKLGPNASTDLKVGDTGFGVHGASQTDPKNGAFABYARVYPP 112
Db 62 -----FDVAGEIVEIGSDIE-NWQVGRWA-----TNIK-GASAEYALIP 102
Qy 113 LFYKSNLTHSTADEISSEGPVKNFESASLPVSLTTAGVSLCHLGSXMEWHPSTPQHTHP 172
Db 103 LLFP--LPKSV-----SYEEGAALAMTFAHLSL-----FDRGLQGET 141
Qy 173 LLWGGATAVGQQLIQVAKHINAYTKIVTVASKKHEKLLKSYGADVDVHDAGVIEQIK 232
Db 142 VLIYGGGAVGHAAIQIAKRAGA-TVITTAGDREKGEIAKAGADQVIFYKE----- 192
Qy 233 SKYPNLQHVDAVGSSEDSIPEAYKVTDADSLPATLLEVVWPTIE-SIPEEIRKD 284
Db 193 -----ESVYDAVGK-----ATNGQGVPLIDMSUSENAQD 223
```



OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
 RN NCBI\_TaxID=1282;  
 RC STRAIN=A3(2) / M145;  
 RX MEDLINE=21996410; PubMed=12000953;  
 RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,  
 Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,  
 Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
 Cronin A., Fraser A., Goble A., Hidaigo J., Hornsby T., Howarth S.,  
 Huang C.-H., Kleser T., Larke L., Murphy L., Oliver K., O'Neill S.,  
 Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,  
 Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
 Warren T., Wietzorrek A., Woodward J., Barrall B.G., Parkhill J.,  
 Hopwood D.A.;  
 RA "Complete genome sequence of the model actinomycete Streptomyces  
 coelicolor A3(2).";  
 RT Nature 417:141-147(2002).  
 RL ENBL; AL939105; CAB59716.1; --  
 DR InterPro; IPR002085; Adh\_zn\_family.  
 DR Pfam; PF00107; adh\_zinc; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 329 AA; 34155 MW; 5DCAD4PB174FD042 CRC64;  
 Query Match 9.2%; Score 176; DB 16; Length 329;  
 Best Local Similarity 28.4%; Pred. No. 2.7e-05;  
 Matches 66; Conservative 28; Mismatches 78; Indels 60; Gaps 10

QY 16 KAVVK-----TDVSPPELKEGTALVKEAVAGNPTDM-----KHIAVKIG--- 55  
 DB 2 KAVQERFPPDLSRLRDVDRPHAGAGQVLRVHAAVNPYDMHMLRGDPYVARMLGGM 61  
 QY 56 ---PEGIIGCDIAGTVVKLGPNASTDLKVGDTGFGFVHGASQTDPKNGAPFAEYARVYPP 112  
 DB 62 LTRPKCPVAGLDAAGVVERVGADV-RGFGPGRVLGFC-----PCFAEYACTTAP 111  
 QY 113 LFKYKSLNLTSTADETISEGPKNFESAASLPVSLTTA--GVSLCHLGSMEHHPSTPQHT 170  
 DB 112 MLV-----PVPEG--LTFEQAAALPMAAVTALRGITRVGVRGQR----- 150  
 QY 171 HPLLIWGGATAVGQQLIQVAKHINAYTKIVTVASKKHEKLKLSYGADVDY 222  
 DB 151 --VLVNGAGGGVGTFAVQIAAVLDA--EVTGVCSSAANTDLVRSLSGAHVLDY 198

RESULT 29  
 Q97V32 PRELIMINARY; PRT; 331 AA.  
 AC Q97V32; (TREMBlrel. 18, Created)  
 DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
 DE Alcohol dehydrogenase (Zn containing) (adh-12) (EC 1.1.1.1).  
 GN ADH-12 OR SSO280.  
 OS Sulfolobus solfataricus.  
 OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;  
 OC Sulfolobus.  
 OX NCBI\_TaxID=2287;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 35092 / DSM 1617 / P2;  
 RX MEDLINE=21332296; PubMed=11427726;  
 RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,  
 Aways M.J., Chan-Weher C.C.-Y., Clausen I.G., Curtis B.A.,  
 De Moore A., Etrauso G., Fletcher C., Gordon P.M.K.,  
 Hoikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,  
 Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,  
 Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,  
 Garrett R.A., Ragan M.A., Sengen C.W., Van der Oost J.;  
 RA "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).  
 RL ENBL; AE006874; AAK42913.1; --  
 DR InterPro; IPR002085; Adh\_zn\_family.  
 DR Pfam; PF00107; adh\_zinc; 1.  
 KW Oxidoreductase; Complete proteome.

OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
 RN NCBI\_TaxID=1282;  
 RC STRAIN=A3(2) / M145;  
 RX MEDLINE=21996410; PubMed=12000953;  
 RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,  
 Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,  
 Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
 Cronin A., Fraser A., Goble A., Hidaigo J., Hornsby T., Howarth S.,  
 Huang C.-H., Kleser T., Larke L., Murphy L., Oliver K., O'Neill S.,  
 Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,  
 Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
 Warren T., Wietzorrek A., Woodward J., Barrall B.G., Parkhill J.,  
 Hopwood D.A.;  
 RA "Complete genome sequence of the model actinomycete Streptomyces  
 coelicolor A3(2).";  
 RT Nature 417:141-147(2002).  
 RL ENBL; AL939105; CAB59716.1; --  
 DR InterPro; IPR002085; Adh\_zn\_family.  
 DR Pfam; PF00107; adh\_zinc; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 329 AA; 34155 MW; 5DCAD4PB174FD042 CRC64;  
 Query Match 9.2%; Score 176; DB 16; Length 329;  
 Best Local Similarity 28.4%; Pred. No. 2.7e-05;  
 Matches 66; Conservative 28; Mismatches 78; Indels 60; Gaps 10

QY 16 KAVVK-----TDVSPPELKEGTALVKEAVAGNPTDM-----KHIAVKIG--- 55  
 DB 2 KAVQERFPPDLSRLRDVDRPHAGAGQVLRVHAAVNPYDMHMLRGDPYVARMLGGM 61  
 QY 56 ---PEGIIGCDIAGTVVKLGPNASTDLKVGDTGFGFVHGASQTDPKNGAPFAEYARVYPP 112  
 DB 62 LTRPKCPVAGLDAAGVVERVGADV-RGFGPGRVLGFC-----PCFAEYACTTAP 111  
 QY 113 LFKYKSLNLTSTADETISEGPKNFESAASLPVSLTTA--GVSLCHLGSMEHHPSTPQHT 170  
 DB 112 MLV-----PVPEG--LTFEQAAALPMAAVTALRGITRVGVRGQR----- 150  
 QY 171 HPLLIWGGATAVGQQLIQVAKHINAYTKIVTVASKKHEKLKLSYGADVDY 222  
 DB 151 --VLVNGAGGGVGTFAVQIAAVLDA--EVTGVCSSAANTDLVRSLSGAHVLDY 198

RESULT 29  
 Q97V32 PRELIMINARY; PRT; 331 AA.  
 AC Q97V32; (TREMBlrel. 18, Created)  
 DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
 DE Alcohol dehydrogenase (Zn containing) (adh-12) (EC 1.1.1.1).  
 GN ADH-12 OR SSO280.  
 OS Sulfolobus solfataricus.  
 OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;  
 OC Sulfolobus.  
 OX NCBI\_TaxID=2287;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 35092 / DSM 1617 / P2;  
 RX MEDLINE=21332296; PubMed=11427726;  
 RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,  
 Aways M.J., Chan-Weher C.C.-Y., Clausen I.G., Curtis B.A.,  
 De Moore A., Etrauso G., Fletcher C., Gordon P.M.K.,  
 Hoikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,  
 Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,  
 Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,  
 Garrett R.A., Ragan M.A., Sengen C.W., Van der Oost J.;  
 RA "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).  
 RL ENBL; AE006874; AAK42913.1; --  
 DR InterPro; IPR002085; Adh\_zn\_family.  
 DR Pfam; PF00107; adh\_zinc; 1.  
 KW Oxidoreductase; Complete proteome.

OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
 RN NCBI\_TaxID=1282;  
 RC STRAIN=A3(2) / M145;  
 RX MEDLINE=21996410; PubMed=12000953;  
 RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,  
 Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,  
 Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
 Cronin A., Fraser A., Goble A., Hidaigo J., Hornsby T., Howarth S.,  
 Huang C.-H., Kleser T., Larke L., Murphy L., Oliver K., O'Neill S.,  
 Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,  
 Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
 Warren T., Wietzorrek A., Woodward J., Barrall B.G., Parkhill J.,  
 Hopwood D.A.;  
 RA "Complete genome sequence of the model actinomycete Streptomyces  
 coelicolor A3(2).";  
 RT Nature 417:141-147(2002).  
 RL ENBL; AL939105; CAB59716.1; --  
 DR InterPro; IPR002085; Adh\_zn\_family.  
 DR Pfam; PF00107; adh\_zinc; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 329 AA; 34155 MW; 5DCAD4PB174FD042 CRC64;  
 Query Match 9.2%; Score 176; DB 16; Length 329;  
 Best Local Similarity 28.4%; Pred. No. 2.7e-05;  
 Matches 66; Conservative 28; Mismatches 78; Indels 60; Gaps 10

QY 16 KAVVK-----TDVSPPELKEGTALVKEAVAGNPTDM-----KHIAVKIG--- 55  
 DB 2 KAVQERFPPDLSRLRDVDRPHAGAGQVLRVHAAVNPYDMHMLRGDPYVARMLGGM 61  
 QY 56 ---PEGIIGCDIAGTVVKLGPNASTDLKVGDTGFGFVHGASQTDPKNGAPFAEYARVYPP 112  
 DB 62 LTRPKCPVAGLDAAGVVERVGADV-RGFGPGRVLGFC-----PCFAEYACTTAP 111  
 QY 113 LFKYKSLNLTSTADETISEGPKNFESAASLPVSLTTA--GVSLCHLGSMEHHPSTPQHT 170  
 DB 112 MLV-----PVPEG--LTFEQAAALPMAAVTALRGITRVGVRGQR----- 150  
 QY 171 HPLLIWGGATAVGQQLIQVAKHINAYTKIVTVASKKHEKLKLSYGADVDY 222  
 DB 151 --VLVNGAGGGVGTFAVQIAAVLDA--EVTGVCSSAANTDLVRSLSGAHVLDY 198

RESULT 29  
 Q97V32 PRELIMINARY; PRT; 3

SQ SEQUENCE 331 AA; 36851 MW; A3A1F1F4EE177A61 CRC64;  
Query Match 9.2%; Score 176; DB 17; Length 331;  
Best Local Similarity 25.5%; Pred. No. 2.7e-05;  
Matches 80; Conservative 51; Mismatches 107; Indels 76; Gaps 16;  
QY 21 TDVSVPELKEGTALVKVEAVAGNPTDWHI--AYKIGPEGSILGCDIAGTVVVKLGPNASTD 79  
DB 18 TDYKNPEIGNEHLVRKLAGVNPVDYVYVERLKNPPIPHIPGVFSGEVVKVGDHVK- 76  
QY 80 LKVGD--TGFGFV-----HGASQTDPKNGAFAYARVYPPLPYKSNLT 120  
DB 77 VSVGDRVTIYGRIFDGTCDMAGVETVCRNGRIGVDANGWAEYVAVBEKYVFKLPNE 136  
QY 121 HSTADEISEGPKVKNFESAASLPVSLTTAGVSLCHLGSKMEWHSTPQHTPLLIWGAT 180  
DB 137 YT-----WEMGSSLTVAAALTA-----YHALKEAQLSPS-----QTLVVFASG 174  
QY 181 AVGOOLIQVAKHINAYTKIVTVASKGHEKLKSYGADDVFDYHDAGVIEQIK-----SKYP 236  
DB 175 NTGMLFVLQKKFGA--KVIAVRK---SWLRGYGADFVDYNE--VEEKKEITNGKMA 227  
QY 237 NLQHVDAVGS--DSIPEAYKVTADSLP-ATLLEVVPMTIESIPBEIRKDNKVIDITLL 293  
DB 228 DV--VINSLGQFWDKFSVLGVRGKLVTFGTLL-----GGNVKVDLSQL 270  
QY 294 YRASQREILLGATR 307  
DB 271 Y--SKHISILGNR 282

RESULT 30  
Q8UKD5 PRELIMINARY; PRT; 358 AA.  
ID Q8UKD5; AC Q8UKD5;  
DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
DE Zinc-binding dehydrogenase.  
GN ATU5188 OR AGR\_FAT\_262.  
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).  
OG Plasmid AT.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
OC Rhizobiaceae; Rhizobium.  
OX NCBI\_TaxID=176299;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21608550; PubMed=11743193;  
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,  
RA Okura Y.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,  
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,  
RA Chapman P., Clendinning J., Deatherage G., Gillet W., Grant C.,  
RA Kutayavin T., Levy R., Li M.-J., McClelland E., Palmieri A.,  
RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,  
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,  
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,  
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,  
RA Nester E.W.;  
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens  
C58.";  
RT Science 294:2317-2323 (2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21608551; PubMed=11743194;  
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,  
RA Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,  
RA Houmieu K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,  
RA Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,  
RA Flanagan C., Crowell C., Gursen J., Lomo C., Sear C., Strub G.,  
RA Cielo C., Slater S.;  
RT "Genome sequence of the plant pathogen and biotechnology agent  
Agrobacterium tumefaciens C58.";  
RT Science 294:2323-2328 (2001).

DR EMBL; AE008941; AAL45877.1; ALT\_INIT.  
DR EMBL; AE007889; AAK90558.1; -  
DR InterPro; IPR002085; Adh\_zn\_family.  
DR Pfam; PF00107; adh\_zinc; 1.  
KW Plasmid; Complete proteome.  
SQ SEQUENCE 358 AA; 38909 MW; 929D84D9D9ED11C5 CRC64;  
Query Match 9.1%; Score 174.5; DB 16; Length 358;  
Best Local Similarity 25.2%; Pred. No. 4e-05;  
Matches 98; Conservative 50; Mismatches 148; Indels 93; Gaps 18;  
QY 2 SVPTTQ-----KAVIIE---GDKAVVKTIDVSPPELKEGTALVKVEAVAGNPTDWHI--KHIA 51  
DB 21 SLFSIQEPEPMKAMSLKSGFGPEAFDLVEVPKPLPKAGQVLRVHATSIINPLDYQVRRGD 80  
QY 52 YK-IGPEGSILGCDIAGTVVVKLGPNASTDLKVGDTGTFGVHGAASQTDPKNGAFAYARVY 110  
DB 81 YRDLVPLPAITGHDVSGVVEATGPGV-TWPAFGD-----VWYTPQIFDGFSGSYAEYHVN 135  
QY 111 PPLFYK--SNLTHSTADEISEGPKVKNFESAASLPVSLTTAGVSLCHLGSKMEWHSTPQ 168  
DB 136 ENTIIGRKPSSLTHL-----EASLSLVGGTAMEALVSRALRV----- 173  
QY 169 HTHPLLIWGATAVGOOLIQVAKHINAYTKIVTVASKGHEKLKSYGADDVFDYHDAGVI 228  
DB 174 -GESSILHGGAGGVGHVAIQVAKAIGA--KVYTTVREENFEFAPARSVGADVVDYRKEDYV 230  
QY 229 EQIKSKYVNL--OHVIDAVGSEDSIPEAYKVTADSLPATLLEVVPMTIESIPBEIRKDNV 286  
DB 231 AAIMRETEGLGVDVDFDTLGE-----TLSHSPKVLQAF--GRVVSIVDIARPN- 278  
QY 287 KIDITLLYRASGOEILLGATRPPASPEYHEATVKFVKFINPHLNNGDIHHMNIKFVSN-- 344  
DB 279 -----LIEAWGR-----NASYH-----FVFTRONQKLNELNVLVERGQL 313  
QY 345 -----GLDDVPALTEGKEGKNKVK 365  
DB 314 RPHVGAVYSIADLP-LAHALLEKPNNGLR 341

Search completed: January 10, 2004, 10:07:16  
Job time : 95 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 10, 2004, 05:10:48 ; Search time 4369 Seconds  
(without alignments)  
10421.702 Million cell updates/sec

Title: US-10-081-644-1  
Perfect score: 1113  
Sequence: 1 atgcaggttccaccactca.....agtatgtgcaggattataa 1113

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0  
Searched: 2888711 seqs, 2045481386 residues  
Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 90 summaries

Database : GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_on.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_om.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pl.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*
- 29: em\_vi.\*
- 30: em\_htg\_hum.\*
- 31: em\_htg\_inv.\*
- 32: em\_htg\_other.\*
- 33: em\_htg\_mus.\*
- 34: em\_htg\_pln.\*
- 35: em\_htg\_rod.\*
- 36: em\_htg\_mam.\*
- 37: em\_htg\_vrt.\*
- 38: em\_sy.\*
- 39: em\_htgo\_hum.\*
- 40: em\_htgo\_mus.\*
- 41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1113	100.0	1113	6	AX524811	Sequence
2	1113	100.0	1113	6	BD174421	Novel eno
3	502.2	45.1	509	6	AX524824	Sequence
4	502.2	45.1	509	6	BD174428	Novel eno
5	359	32.3	1134	6	AX524815	Sequence
6	359	32.3	1134	6	BD174423	Novel eno
C 7	359	32.3	29634	8	YSL9122	Saccharomyc
8	357.8	32.1	1145	6	AX524813	Sequence
9	357.8	32.1	1145	6	BD174422	Novel eno
10	357.8	32.1	1631	6	AX536730	Sequence
C 11	357.8	32.1	1944	8	SCYNL134C	271410 S.cerevisia
C 12	357.8	32.1	43481	8	SCCHXIV43	246843 S.cerevisia
C 13	347.8	31.2	316613	8	SCCHRII	X59720 S.cerevisia
14	345.2	31.0	1122	6	AX524817	Sequence
15	345.2	31.0	1122	6	BD174424	Novel eno
16	252.4	22.7	1083	11	CNS06FEH	AL396393 17 end of
C 17	103.6	9.3	5355	8	YSCDC91A	AL402417 17 end of
C 18	103.2	9.3	5355	8	AX536956	Saccharomyc
19	71	6.4	1086	6	AX536956	Sequence
20	53.2	4.8	1506	8	CCTOXD	X92391 C.carbonum
C 21	48.6	4.4	212955	2	AC136676	Rattus no
C 22	48.4	4.3	2000	6	AX655393	Sequence
23	48.2	4.3	109375	8	AC003672	Arabidops
24	47.6	4.3	84550	3	PFMAL1P2_3	Continuation (4 of
25	47.4	4.3	13844	1	AE007561	AE007561 Clostridi
26	47.4	4.3	92357	9	HSJ106901	AL049633 Human DNA
27	47.4	4.3	119082	2	AC008057	Human DNA
C 28	47.4	4.3	169494	9	AC007630	AC008057 Homo sapi
C 29	47.4	4.3	254449	3	AB014817	AC007630 Homo sapi
30	47	4.2	84118	2	AC096684	Plasmodiu
31	47	4.2	179553	2	AC024253	AC096684 Takifugu
32	46.4	4.2	154071	3	AC115598	AC024253 Homo sapi
C 33	46.2	4.2	5537	3	AP465308	AC115598 Dictyoste
34	46	4.1	110000	2	PFMAL8P1_09	AF465308 Dictyoste
C 35	45.8	4.1	14422	3	AP466146	Continuation (10 o
36	45.6	4.1	341050	3	PFAS29357	AF466146 Melipona
C 37	45.4	4.1	9553	5	AP134321	AL929357 Plasmodiu
38	45.4	4.1	303414	1	AE015938	AF134321 Diessostic
39	44.8	4.0	2922	5	LFL308116	AE015938 Clostridi
40	44.8	4.0	164487	2	AL954708	AE015938 Lampetra
C 41	44.8	4.0	242302	2	AC131200	AL954708 Danio rer
C 42	44.6	4.0	11327	1	AE006787	AC131200 Rattus no
C 43	44.6	4.0	163243	9	AL513304	AE006787 Sulfolobu
44	44.2	4.0	74154	2	AC096997	AL513304 Human DNA
C 45	44.2	4.0	186935	2	AC022322	AC096997 Takifugu
C 46	44.2	4.0	236263	2	AC098600	AC022322 Homo sapi
47	44	4.0	244845	2	AC126846	AC098600 Rattus no
48	43.8	3.9	157695	9	AL136442	AC126846 Rattus no
49	43.8	3.9	213544	2	BX510939	AL136442 Human DNA
C 50	43.8	3.9	221756	2	AC094893	BX510939 Danio rer
C 51	43.8	3.9	232974	2	AC114193	AC094893 Rattus no
C 52	43.8	3.9	253001	3	AB014834	AC114193 Rattus no
53	43.8	3.9	299650	1	AP004603	AB014834 Plasmodiu
C 54	43.8	3.9	325350	1	AP004823	AP004603 Oceanobac
C 55	43.6	3.9	242516	2	AC106523	AP004823 Staphyloc
56	43.6	3.9	243203	2	AC090437	AC106523 Rattus no
C 57	43.6	3.9	250541	1	AE017004	AC090437 Mus muscu
C 58	43.6	3.9	294769	2	AC098907	AE017004 Bacillus
C 59	43.4	3.9	110000	2	PFMAL7P1_05	AC098907 Rattus no
C 60	43.2	3.9	94434	5	AL929250	Continuation (6 of
C 61	43.2	3.9	110000	2	PFMAL7P1_08	AL929250 Zebrafish
62	43.2	3.9	184090	10	AC122304	Continuation (9 of
C 63	43.2	3.9	191613	10	AC132408	AC122304 Mus muscu
C 64	43.2	3.9	243490	2	AC131536	AC132408 Mus muscu
C 65	43.2	3.9	249995	3	AE014840	AC131536 Rattus no
						AE014840 Plasmodiu

66	43.2	3.9	313096	2	AC131879	AC131879 Rattus no
67	42.8	3.8	2000	6	AX655393	AX655393 Sequence
68	42.8	3.8	116696	3	PFMAL3P3	Z98547 Plasmodium
69	42.6	3.8	4111	10	MMU278580	AJ278580 Mus muscu
70	42.6	3.8	125026	10	AC091616	AC091616 Rattus no
71	42.6	3.8	178273	2	AC005308	AC005308 Plasmodiu
72	42.6	3.8	183420	2	AC116417	AC116417 Mus muscu
73	42.6	3.8	185814	2	AC116888	AC116888 Mus muscu
74	42.6	3.8	198550	2	AC140782	AC140782 Mus muscu
75	42.6	3.8	219955	2	AC127414	AC127414 Mus muscu
76	42.6	3.8	241327	2	AC127126	AC127126 Rattus no
77	42.6	3.8	251075	2	AC132995	AC132995 Rattus no
78	42.6	3.8	269050	1	AP000987	AP000987 Sulfolobu
79	42.6	3.8	349751	3	PFMAL4P3	AL035476 Plasmodiu
80	42.4	3.8	171730	2	AL929558	AL929558 Danio rer
81	42.4	3.8	291490	1	AB017030	AB017030 Bacillus
82	42.2	3.8	206836	2	AC129047	AC129047 Rattus no
83	42.2	3.8	222603	2	AC125755	AC125755 Rattus no
84	42.2	3.8	234097	2	AC120493	AC120493 Rattus no
85	42.2	3.8	248793	2	AC103080	AC103080 Rattus no
86	42.2	3.8	301050	1	AP003130	AP003130 Staphyloc
87	42.2	3.8	333321	3	AC116986	AC116986 Dictyoste
88	42.2	3.8	343590	1	AP003359	AP003359 Staphyloc
89	42	3.8	96573	2	AC136672	AC136672 Rattus no
90	42	3.8	110000	2	AC111624_2	Continuation (3 of
ALIGNMENTS						
RESULT 1	AX524811	AX524811	1113 bp	DNA	linear	PAT 21-NOV-2002
LOCUS	Sequence 1 from Patent EP1236796.					
DEFINITION	AX524811					
ACCESSION	AX524811.1	GI:25169907				
VERSION						
KEYWORDS	Kluyveromyces lactis					
SOURCE	Kluyveromyces lactis					
ORGANISM	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Kluyveromyces.					
REFERENCE	1					
AUTHORS	Yamamoto,H. and Kimoto,N.					
TITLE	Novel enzyme reductions isolated from kluyveromyces lactis, methods for producing same, and methods for selectively reducing a carbon-carbon double bond of an alpha, beta-unsaturated ketone using the reductions					
JOURNAL	Patent: EP 1236796-A 1 04-SEP-2002; Daicel Chemical Industries, Ltd. (JP)					
FEATURES	Location/Qualifiers					
source	1..1113					
	/organism="Kluyveromyces lactis"					
	/mol_type="genomic DNA"					
	/db_xref="taxon:28985"					
CDS	1..1113					
	/note="unnamed protein product"					
	/codon_start=1					
	/protein_id="CAD57033.1"					
	/db_xref="GI:25169908"					
	/translation="MSVPTQKAVIIEGDKAVVKTDSVPPELKEGTALVKVEAVGNP TDWKHAIYKIGPEGSILGCDIAGTVKVGPNASTDLKVGDTGFGVHGASQDTPKNGA FAEVARYVPLFYKSNLTHSTADEISEGPVKNFESAASLPVSLTTAGVSLCHHLSKM EWPSTPQTHPLLINGGATAGVQQLIQVAKHINAYKIVTVASKHEKLLKSYGADD VFDHDAVIEQIKSKYPTNLOHVDAVGSSEDSIPEAYKVTADSLPATLLEVPMTIES IPEIRKNDKIDITILYRASGOEILGATRFPAPEYHEATVKFKVFNPHNLNGDI HHMNKVFNSGLDVPALTEGIEGKKNKYVARL"					
BASE COUNT	343 a	210 c	234 g	326 t		
ORIGIN						
Query Match	100.0%;	Score 1113;	DB 6;	Length 1113;		
Best Local Similarity	100.0%;	Pred. No. 6e-238;				
Matches 1113;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;		



JOURNAL Patent: EP 1236796-A 14 04-SEP-2002;  
Daicel Chemical Industries, Ltd. (JP)  
FEATURES Location/Qualifiers  
source 1..509  
/organism="Kluyveromyces lactis"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:28985"  
BASE COUNT 136 a 108 g 156 t 5 others  
ORIGIN

Query Match 45.1%; Score 502.2; DB 6; Length 509;  
Best Local Similarity 98.4%; Pred. No. 1.3e-101;  
Matches 501; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 160 ATTGGTCCAGAGGTTCAATCTAGGATGTGACATTCTGGTACAGTTGTCAAACTTGA 219  
DB 1 ATTGGTCCAGAGGTTCAATCTAGGATGTGACATTCTGGTACAGTTGTCAAACTTGA 60

QY 220 CCAAAATGCTAGTACTGACTTGAAGTTGGAGATACCGGTTTCGGTTTGTTCACGGTGT 279  
DB 61 CCAAAATGCTAGTACTGACTTGAAGTTGGAGATACCGGTTTCGGTTTGTTCACGGTGT 120

QY 280 TCCCAACACAGATCCTAAATATGCTGAATATGCCAGGGTTTATCCACCTTTG 339  
DB 121 TCCCAACACAGATCCTAAATATGCTGAATATGCCAGGGTTTATCCACCTTTG 180

QY 340 TTTTACAGAGTAATTAACCTCACTCACTGCTGATGAAATTTCTGAAGCCCTGTGAAG 399  
DB 181 TTTTACAGAGTAATTAACCTCACTCACTGCTGATGAAATTTCTGAAGCCCTGTGAAG 240

QY 400 AACTTCGAATCTGCTGCATCAATGCCAGTTTGGTTGACAACTGCTGGTGTAGTTGTGT 459  
DB 241 AACTTCGAATCTGCTGCATCAATGCCAGTTTGGTTGACAACTGCTGGTGTAGTTGTGT 300

QY 460 CATCACTTGGGCTCAAAAATGGAATGGCAACCCATCTACCCCGCAACATATC 519  
DB 301 CATCACTTGGGCTCAAAAATGGAATGGCAACCCATCTACCCCGCAACATATC 360

QY 520 TTGATTTGGGCTGGTGTACAGAGTGGTCAACAACTAATCAAGTTGCCAAACATATC 579  
DB 361 TTGATTTGGGCTGGTGTACAGAGTGGTCAACAACTAATCAAGTTGCCAAACATATC 420

QY 580 AATGCTTATACAAAGTGTGAAGTTGCTTCTAAAGAGCATGAAAGCTTTTAAAGTCT 639  
DB 421 AATGCTTATACAAAGTGTGAAGTTGCTTCTAAAGAGCATGAAAGCTTTTAAAGTCT 480

QY 640 TATGGTGTGATGATGCTTCTTGCATATCA 668  
DB 481 TATGGTGTGATGATGCTTCTTGCATATCA 509

RESULT 4  
BD174428  
LOCUS 509 bp DNA linear PAT 18-MAR-2003  
DEFINITION Novel enzyme reductase, process for producing the same, and method for selectively reducing carbon-carbon double bonds of alpha, beta-unsaturated ketone using the same.  
BD174428  
VERSION BD174428.1 GI:29120116  
KEYWORDS JP 2002247987-A/8.  
SOURCE Kluyveromyces lactis  
ORGANISM Kluyveromyces lactis  
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Kluyveromyces.  
Yamamoto.H. and Kimoto, K.  
TITLE Novel enzyme reductase, process for producing the same, and method for selectively reducing carbon-carbon double bonds of alpha, beta-unsaturated ketone using the same  
JOURNAL Patent: JP 2002247987-A 8 03-SEP-2002;  
DAICEL CHEMICAL INDUSTRIES LTD  
COMMENT OS Kluyveromyces lactis  
PN JP 2002247987-A/8

PD 03-SEP-2002  
PF 23-FEB-2001 JP 2001049363  
PI HIROAKI YAMAMOTO, KUNIHICO KIMOTO  
PC C12N15/09, C12N1/15, C12N1/19, C12N1/21, C12N5/10, C12N9/02, C12P7/26//  
PC (C12N15/09, C12R1:645), C12N15/00, C12N5/00, (C12N15/00, C12R1:645)  
CC Novel enzyme reductase, process for producing the same, and CC method for  
CC selectively reducing carbon-carbon double bonds of alpha, beta-unsaturated  
CC ketone using the same  
FH Key Location/Qualifiers  
FT source 1..509  
FT /organism="Kluyveromyces lactis".  
FEATURES Location/Qualifiers  
source 1..509  
/organism="Kluyveromyces lactis"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:28985"  
BASE COUNT 136 a 104 c 108 g 156 t 5 others  
ORIGIN

Query Match 45.1%; Score 502.2; DB 6; Length 509;  
Best Local Similarity 98.4%; Pred. No. 1.3e-101;  
Matches 501; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 160 ATTGGTCCAGAGGTTCAATCTAGGATGTGACATTCTGGTACAGTTGTCAAACTTGA 219  
DB 1 ATTGGTCCAGAGGTTCAATCTAGGATGTGACATTCTGGTACAGTTGTCAAACTTGA 60

QY 220 CCAAAATGCTAGTACTGACTTGAAGTTGGAGATACCGGTTTCGGTTTGTTCACGGTGT 279  
DB 61 CCAAAATGCTAGTACTGACTTGAAGTTGGAGATACCGGTTTCGGTTTGTTCACGGTGT 120

QY 280 TCCCAACACAGATCCTAAATATGCTGAATATGCCAGGGTTTATCCACCTTTG 339  
DB 121 TCCCAACACAGATCCTAAATATGCTGAATATGCCAGGGTTTATCCACCTTTG 180

QY 340 TTTTACAGAGTAATTAACCTCACTCACTGCTGATGAAATTTCTGAAGCCCTGTGAAG 399  
DB 181 TTTTACAGAGTAATTAACCTCACTCACTGCTGATGAAATTTCTGAAGCCCTGTGAAG 240

QY 400 AACTTCGAATCTGCTGCATCAATGCCAGTTTGGTTGACAACTGCTGGTGTAGTTGTGT 459  
DB 241 AACTTCGAATCTGCTGCATCAATGCCAGTTTGGTTGACAACTGCTGGTGTAGTTGTGT 300

QY 460 CATCACTTGGGCTCAAAAATGGAATGGCAACCCATCTACCCCGCAACATATC 519  
DB 301 CATCACTTGGGCTCAAAAATGGAATGGCAACCCATCTACCCCGCAACATATC 360

QY 520 TTGATTTGGGCTGGTGTACAGAGTGGTCAACAACTAATCAAGTTGCCAAACATATC 579  
DB 361 TTGATTTGGGCTGGTGTACAGAGTGGTCAACAACTAATCAAGTTGCCAAACATATC 420

QY 580 AATGCTTATACAAAGTGTGAAGTTGCTTCTAAAGAGCATGAAAGCTTTTAAAGTCT 639  
DB 421 AATGCTTATACAAAGTGTGAAGTTGCTTCTAAAGAGCATGAAAGCTTTTAAAGTCT 480

QY 640 TATGGTGTGATGATGCTTCTTGCATATCA 668  
DB 481 TATGGTGTGATGATGCTTCTTGCATATCA 509

RESULT 5  
AX524815  
LOCUS 1134 bp DNA linear PAT 21-NOV-2002  
DEFINITION Sequence 5 from Patent EP1236796.  
AX524815  
ACCESSION AX524815  
VERSION AX524815.1 GI:25169911  
KEYWORDS Saccharomyces cerevisiae (baker's yeast)  
SOURCE Saccharomyces cerevisiae  
ORGANISM Saccharomyces cerevisiae

Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

1  
Yamamoto, H. and Kimoto, N.  
Novel enone reductases isolated from *Kluyveromyces lactis*, methods  
for producing same, and methods for selectively reducing a  
carbon-carbon double bond of an alpha, beta-unsaturated ketone  
using the reductases  
Patent: JP 1236796-A 5 04-SEP-2002;  
Daicel Chemical Industries, Ltd. (JP)

JOURNAL  
Location/Qualifiers  
1. .1134  
/organism="Saccharomyces cerevisiae"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:4932"

CDS  
1. .1134  
/notes="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAD57035.1"  
/db\_xref="GI:25169912"  
/translation="MAQVAIPEIMKAVVIEDKAVKGIPIPELEBGFVLIKTLAVA  
GNPTDAHIDYKIGPGSILGCDAGOI VKLGPVNPDPISGDIYVGFTHSSVRP  
SNGAPAEYSALSTVAVKSNELKFLGEDVLPAGVRSLEGVATIPVSLTGLVITY  
NLGDLKWEESTPORKPILLNGATVQSLIOLANKUNGFTKIIIVASRKHKLK  
EYGADELFDYHDIDVBOIKHNNISYLVDCVANQDTLOQVYKCAADKQDATIVELK  
NLTEBNYKKNRQNVTDIIRLYSIGGHEVPFGNITLPADSEARAAIKFIKFINPK  
INDQIRHIPVRVYKGLCDVPHILKDIKYKNSGEKLVAVLN"

BASE COUNT 351 a 225 c 245 g 313 t

ORIGIN  
Query Match 32.3%; Score 359; DB 6; Length 1134;  
Best Local Similarity 58.9%; Pred. No. 1.1e-69;  
Matches 658; Conservative 0; Mismatches 450; Indels 9; Gaps 2;

Qy 5 CAGTTCACCAACCAACCAACCGCTCATCTGTAAGGTGACAAAGCTGTGTTAAACAG 64  
Db 14 CAATTCAGAAACCAACCAACCGCTGTCTATTAAGACGCGTGTGTTAAAGAGG 73

Qy 65 ATGTCCTCAGTTCACAGAAATTAAGGAGGTACAGCGCTTGGTGAAGCTTGAGCGTCTGCTG 124  
Db 74 GCATTCCTCATCTGTAATTAAGGAGGATTCGTATTAAGTAAACACTGCTGTGCTG 133

Qy 125 GTAACCAACCTGATGGAAGCATATGCTTATAAGATTGGTCCAGAGGTTCAATTCAG 184  
Db 134 GTAACCAACCTGATGGAAGCATATGCTTATAAGATTGGTCCAGAGGTTCAATTCAG 193

Qy 185 GATGTGACATTCGCTGATCAGTTGCTCAACCTGACCAACCAATGCTAGTACT---GACTGA 241  
Db 194 GATGTGATGCTGCTGGCCAAATTTGTCAAAATGGGCCAGCTGTCAATCCTAAAGACTTTT 253

Qy 242 AGGTTGGAGATACCGGTTTTCGGTTTGTTCACGGTGTCTCCAAACAGATCTCAAAATG 301  
Db 254 CTATCGGTGATATATTAATGAGTTTCAATTCAGGATCTTCGTAAGGTTTCCTTCCATG 313

Qy 302 GTGCATTTGCTGAATATGCCAGGTTTATCCACCTTGTGTTTACAA-----GAGTAACT 355  
Db 314 GTGCTTTGCTGAATATTCCTGATTAATCACTGTTGGTTCGCTACAAATCAACCAATGAC 373

Qy 356 TAACTCACTCACTGCTGATGAATAATTCGTAAGGCCCTGTGAAGAACTTCGAATCTGCTG 415  
Db 374 TCAAAATTTTGGGTGAGGATTTCTTACCTGCGGCCCTGTGAGGTCTTTGGAGAGGTGAG 433

Qy 416 CATCATTCGCCAGTTTCGTTGACAACTGCTGTTGTTAGTTGTCATCACTTGGGCTGCTCA 475  
Db 434 CCATATCCCACTGCTGATGACCAAGCGGCTGTTGTTGACCTATTAACCTTGGGCTG 493

Qy 476 AAATGGATGGCAACCCATCTACCCCGCAACATACTCATCTCAATTAATTTGGGCTGCTG 535  
Db 494 ACCTGAGTGGGACCATCAACCCCAACAAAGAGGCCCATCTTATTATGGGCGGCTG 553

Qy 536 CTACAGCAGTGGGTCAACAACTAATCAAGTTGGCAACATATCAATCAATGCTTATCAAGA 595  
Db 554 CAATGTCAGTAGGTGAGTGGCTGCTCATCCAAATTAGCCAAATTAATGAGTGGCTTCAACAGA 613

Qy 596 TTCTAACTGTTGCTTCTTAAAGCATGAAAGCTTTTAAAGCTTTATGCTGCTGATGATG 655  
Db 614 TCATTGTTGCTGCTTCTCGGAGCAGCAAAACCTTTTGAAGAAATATGTTGCTGATGAAT 673

Qy 656 TCTTTGACTATCATGATGACGCGTATTATGAGCAGATCAATCGAAGTATCCAAACCTGC 715  
Db 674 TATTTGATTATCATGATATTGACGTGTAGAACAAAATTAACACAAAGTACAAACATATCT 733

Qy 716 AACATGTTATTGACGCTGTGGGAAGCAAGATAGTATCCCGAGGCTTATAAAGTCCAG 775  
Db 734 CGTATTTAGTCGACTGTGTCGGAATCAAGATACGCTTCAACAGTGTACAAATGTCGG 793

Qy 776 CAGATAGTCTACCTGCCACATTTATTAGAAGTGGTTTCCAATGACCATTTGAAAGCATTCCTG 835  
Db 794 CCGATAAACAGGATGCTACAAATTTGTAATTAATAAATTTGACAGAGAAACGTCACAAA 853

Qy 836 AAGAAATCAGAAAGATAATATGTTAAATTTGATATTAATCTTTGTTGATCGTATCTGGTC 895  
Db 854 AAGAGAACAGGACACAAACGTTTACTATTGACATAAATAGGCTATATTCAATAGTGGCC 913

Qy 896 AAGAAATCTATTGGGTGCAACAGATTTCTCTAGTCCAGAAATATCATGAAAGCCACAG 955  
Db 914 ATGAAGTACCATTTTGGAAACATTTACTTTACCGCCGACTCAGAAAGCTAGGAAAGCTGCA 973

Qy 956 TTAATTTGCTTAAAGTTTATAAATCCACACCTTAACCAACGGTATATCCATCATATGAATA 1015  
Db 974 TAAATTTATCAATTCATCAATCCAAAGATTAATGATGGCAAAATTCGCCATATTCCAG 1033

Qy 1016 TTAAGTTTTCAGCAACGCTTAGATGATGTCCAGCTCTCAGTGAAGGTATATAAAGAG 1075  
Db 1034 TAAGGGTCTATAAAGACGGGCTTTGTGATGTTCTCATATCTCTAAAGACATCAAAATATG 1093

Qy 1076 GTAAGAACAAAATGTTAAGTATGTTGCCAGGTTATA 1112  
Db 1094 GTAAGAACTCTGTGTAAGAACTCTGTTGCCGTATTAA 1130

RESULT 6  
BD174423 1134 bp DNA linear PAT 18-MAR-2003

LOCUS  
DEFINITION  
Novel enone reductase, process for producing the same, and method  
for selectively reducing carbon-carbon double bonds of alpha,  
beta-unsaturated ketone using the same.

ACCESSION  
BD174423  
VERSION  
BD174423.1 GI:29120111  
KEYWORDS  
JP 2002247987-A/3.  
SOURCE  
Saccharomyces cerevisiae (baker's yeast)  
ORGANISM  
Saccharomycetes  
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

REFERENCE  
1 (bases 1 to 1134)  
Yamamoto, H. and Kimoto, K.  
Novel enone reductase, process for producing the same, and method  
for selectively reducing carbon-carbon double bonds of alpha,  
beta-unsaturated ketone using the same  
Patent: JP 2002247987-A 3 03-SEP-2002;  
DAICEL CHEMICAL INDUSTRIES LTD  
COMMENT  
OS Saccharomycetes cerevisiae (yeast)  
PN JP 2002247987-A/3  
PD 03-SEP-2002  
PF 23-FEB-2001 JP 2001049363  
PI HIRAKI YAMAMOTO KUNIHIRO KIMOTO  
PC C12N15/09, C12N1/15, C12N1/19, C12N1/21, C12N5/10, C12N9/02, C12P7/  
PC 26//  
PC (C12N15/09, C12R1:645), C12N15/00, C12N5/00, (C12N15/00, C12R1:645)  
CC Novel enone reductase, process for producing the same, and CC  
method for  
CC selectively reducing carbon-carbon double  
bonds of alpha, beta-  
unsaturated  
CC ketone using the same  
FH Key Location/Qualifiers

FT	CDS	(1)..(1134).
FEATURES	Location/Qualifiers	
source	1..1134	
	/organism="Saccharomyces cerevisiae"	
	/mol_type="genomic DNA"	
	/db_xref="taxon:4932"	
BASE COUNT	351 a -225 c 245 g 313 t	
ORIGIN		
Query Match	32.3%; Score 359; DB 6; Length 1134;	
Best Local Similarity	58.9%; Pred. No. 1.1e-69;	
Matches	658; Conservative 0; Mismatches 450; Indels 9; Gaps 2;	
Qy	5 CAGTTCACCACTCAAAAGCCGTCATCATTTGAAGGTGACAAAGCTGTTGTTAAACAG 64	
Db	14 CAATTCCAGAAACCATGAAGCGCTGTCGTCATTGAAGCGGTAAAGCGGTGTTAAAGAGG 73	
Qy	65 ATGTCTCAGTTTCAGAAATTAAGAGGGGTACAGCGCTTGGTGAAGGTTTCAGGCTGTTGCTG 124	
Db	74 GCATTCGCATTCCTGAATTTGGAAGAGGATTCGTATGTTAAGACATCGCTGTTGCTG 133	
Qy	125 GTAAACCAACTGATTCGAAGCATATTGCTTTATAAGATTGGTCCAGAGGTTCAATTTCTAG 184	
Db	134 GTAAACCCACTGATTGGGCACACATTGACTACAAGATGGGCTCAAGGATCTATTCTGG 193	
Qy	185 GATGTGACATTCGTGTACAGTTGTCAACTTGGACCAATGCTAGTACT---GACTTGA 241	
Db	194 GATGTGATGCTGTGCCAAATTTGTCAAAATGGGCCCACTGTCTCAATCCTAAAGACTTTT 253	
Qy	242 AGGTGGAGATACCGGTTTCGGTTTGTTCAGCGTCTTCCCAACACAGATCTTAAATG 301	
Db	254 CTATCGTGATATATATTATGGTTCATTCAGGATCTCCGTAGGTTTCTTCCATG 313	
Qy	302 GTGCATTTGCTGAATGCGAGGTTTATCCACCTTTGTTTACAA-----GAGTAAT 355	
Db	314 GTGCTTTTGTGAATATCTGCTATTTCAACTGTGTTGCCCTACAAATCAACCCATGAAC 373	
Qy	356 TAACCTCACTCACTGATGAATTTCTGAAGGCCCTGTGAAGACTTCGAATCTGCTG 415	
Db	374 TCAAAATTTTGGGTGAGGATGTTCTACCTGCCGCCCTGTGAGGTCCTTTGAAGGTGAG 433	
Qy	416 CATCATTTGCCAGTTTCGTGTGACAACTGCTGGTGTAGTTTGTGTCATCACTTTGGGCTCAA 475	
Db	434 CCACATATCCAGTCTCACTGACCACAGCCGGCTTGGTGTGACCTATACTTTGGGCTGG 493	
Qy	476 AAATGGAATGGCAACCATCTACCCCGCAACATCACTATCCATTAATGATTTGGGGGTGGT 535	
Db	494 ACCTGAATGGGAGCCATCAACCCCAAGAAAGAGGCCCATCTTTATTTATGGGGCGGTG 553	
Qy	536 CTACAGCAGTGGGTCAACAACTAACTCCAAGTTTGCACAAACATATCAATCTTACTAAGA 595	
Db	554 CAATCGATGATGCTAGTCGCTCAUCCATTAAGCAATAAATTGAATGGCTTCCACAGA 613	
Qy	596 TTGTAACTGTGTTCTTAAAGAGCATGAAAGCTTTTAAAGTCTTTATGGTCTGATGATG 655	
Db	614 TCATTGTTGGGCTTCTCGAGAGCAGAAACCTTTTGAAGAAATATGCTGCTGATGAAT 673	
Qy	656 TCTTTGACTATCATGATGAGCGGTTATTGACAGATCAATTCGAAGTATCCAAACCTGC 715	
Db	674 TATTTGATTTATCATGATATTGACGTTGGTAGACAAATTTAAACACAAAGTACCAATATCT 733	
Qy	716 AACATGTTATTGACGCTGTGGGAAGGAGATAGTATCCCGAGCGCTTATAAGTTCACAG 775	
Db	734 CGTATTATTGCTGACTGTGTCGGAATCAGATACGCTTCAACAGTGTACAAATGTGGGG 793	
Qy	776 CAGATAGTCTACCTGCCACATTTATTAGAAAGTGGTTCCCAATGACCAATGAAAGCAATTCTGT 835	
Db	794 CCGATAAACAGGATGCTACAAATTTGTGAATTTAAAAAATTTGACAGAAAGAAACGTCACAA 853	
Qy	836 ACGAATATCGAAAGATATGTTAAATTTGATTTACTTTGTTGATCGTGCATCTGGTC 895	
Db	854 AAGAGAACAGGAGACAAACGTTTACTTATGACATTAAGGCTATATTCAATAGGTGGCC 913	

the rightmost cosmid of chromosome 12. The cosmid on the left is 9324. This sequence overlaps the right telomeric sequence, accession number Z7337, starting at nucleotide 28299.

FEATURES	Location/Qualifiers	source
	1..29634	/organism="Saccharomyces cerevisiae" /mol_type="genomic DNA" /strain="S288C (AB972)" /db_xref="taxon:4932" /map="XII"
	51..7937	
	51..7937	
	51..7937	
	51..7937	
	51..7937	
	51..7937	
	51..7937	
	51..7937	
	51..7937	
gene	/evidence=not experimental /product="Yl454wp" /protein_id="AAB64718.1" /db_xref="GI:717075"	
CDS	/translation="MMPPINVLLYKWLIFAVTFELMSCKILLRKLGINITWINLFKLE ICGLSLEDGTGVRKLSVFAVERKLFPIKGLRIDSKSSYNDLHKLPREBERTFIETP EDNGGGFISKILSOLYMLANGVTIILEDOLVNDITIERKFPFLSIDNSKHTKSLRF DSELRKLLWNGQIIADAIFIVTNLLIGIMPLKDLQVGLDGLDGLNINPMLLN LFTKENVDMSNEKLQRLADITKANEEULSDIAQKDDLVYAMEKPVDRIKPLKE MNTVDKQIKDPLTNHPELGNKYISYNVLVSNINENTNFRNEMPGYTIIFPER DSPEKFSINARINILNRKQSHAKQLKIEINPVISFGSTNLFQKRLSNLH AKELNAIPNIKINISLTIDMPVNFIFKISNIPKFTSCPKNKLKENSXKVF LTRRVLFVDEKFLPLINMKFTLDOPKFINDKDLIIGKFSVEMISHSKRYTLGN NLMEKEETQHIYFESHVNVVELDMKQLHIIKQKYEHTILRVDSIAIEBKVQLPDI LCSANADITLMDLSELPTVMVLSLWNLDSQLANEENYKFEYKFPASLONMK AECNSMAKRLKILPSPFMFQQLPFDYIKINIDISHTIAGSVFMRDVFSSV DSQSKDIDGKRLKNTVEKQIQLFGDKTQWNRKIGSNHATMVRSGOLTWFSDN KONPKHSSITADDDISTSDATEVHNLNINLVNDITTSIIGETPEVSELSKTVS KVNSLSKLPDTESSNESDSKILQINHSRGTSVLSMSIFLAVSGHTLNQIFG HCHQNRKSKTQIFUALSESSCKIKWQLKLEINFSEIYSQIILPNG LRTKFEPTSFTIVKNCNTISVSGYFRMMVESPTQPFWMRCINGFKFWKHIDLL KQMKLKNLQWELKLPASITLNDSHWHSIFPHFEMPKIIDSIFTFKSIKQMLYSL AKQDFNKRISDLIKSOKWAKEMODFEAIDNAILKRTGLWAKDGKKRLKRSATDS EIPLPAALNKGKDRDRDPTQFISPEIENAYNTLLANESDSWIKRKYKVERRE FDKGFSLWGFIDYTKLPDKINKVLPFPNPFMLNLIENIDIDIIRPSCGIENIPN FIHDVGKVPNTYEYSIMIPMLHDAKPESEVRHLRDYPLFPFISPLSSSTOSKETIPM RIYGDFTIMDLQSDRELTLFVPLIIPSVTVENTDRIYSLFVPRMTWSAKIFDNLN EINSHNTRTVWGSQPAIQOTWQCLDNFSPKLPDPSVKLGFWDKTRYLFHGKINIV WKRGRFISLGAKSPYMLGSGSAPFIVGFGNVLNKCNEEDPKFLSCADKVFH SINYPFAKPLVNSRSTPMFIIPNODDTNMQRYASFYLLNTTSSKNEKADKEIMGK SFIEKYGKLSGWTLMGILFERLGSINERTFESKLYTLRCLPIYVQDLSKSDS YAGFRSDFIHMSFGLSNSNSAYNAQLSPNGKAFVWPKFSGNFPVRRGFLGLQ SIDPKFGEHLTYTISYHADVSPLFINMYHNADADQILRNKYLEVAEAGLKAHSSHF MDLQRKEVLTYEAGLNVRRVYMKLFLAGDVVQDVIDRTVSGFSLKLYTEEKED AYDIDPNDMSLWDITDQDAFFINPDNYPKIKIMPFAPSPQFAYQKASGYDKYQV DPKTCQIPTPDNRVSHGCTLGHVSLRDLVEKRYTVLAKPREKQEGIRKNSAGV SEENLNDLSKANSVNEABELLQPKQIFQKEAGQEQEPHFDLSNLNLRKTKLK QFBEHFFIIVNLKWNEDASAIKFFKYANLSENFASLASKGLREFEDVIKQREMT DDTTSMEAIPEGDKANTTKQCHSCDDTEFTENLNIIEKNITQLSODIKNKHIFK FVQFITQIOLTSLENPEACVLSPSPFMKLTLEFDANTTNTYMODIFLKRHILGFG NANAFNLKDYQEFRELYFGSSSYGQDKKEQWPMGLGLEPESALKKAVNRNIS ALLHQKLAPEKADSLDKIEDNICGYQVQNVQVNSDEYMLTAKMLKFLIYVEP EDELKYLITKLIIGYDIYDQTRKFVNOLHDSQILAVVEKLELFPKSLDDIDGKL DLNINHNHQRHLRLYLIRKVTNSGNNYINRLVNNIKVNETILHLLDKIDKPFID IYAVKLNFORQIHTLGRKNTVYQMMQIFDLGNGVNHGICLPLTISGNDTVGLAS DVLQVQITFDVDPKPVGGIKVVKVETLTSLTIKLEBDRLNKLFEWLSLKYDNGG DDDDGASSITWQVSSSECKIEFSEDISDFNEMLKRSDYMDVEDKLNSKELCIS YKQKQRLANVTNFVNFPTLRSLNQTLRVTDLLALKVKVLIKVLIKHTRGFIGNKL KRNSKNKIADDTSPKLQTLTYNSPEBELR"	
gene	complement (8124..8197) /gene="t1(AAU) L2"	
tRNA	complement (8124..8197) /gene="t1(AAU) L2"	
repeat_unit	/product="tRNA-ile" /notes="codon recognized: AUU" complement (8357..8685) /note="delta element YLRCDelta27"	
	9682..10596	/gene="YLR455W" 9682..10596 /gene="YLR455W" /codon_start=1 /evidence=not experimental /product="Yl455wp" /protein_id="AAB64719.1" /db_xref="GI:717074"
	9682..10596	
	9682..10596	
	9682..10596	
	9682..10596	
	9682..10596	
	9682..10596	
	9682..10596	
	9682..10596	
	9682..10596	
gene	/translation="NTKDIRTGDVLCKVSGPPMPVPPORLLANDVVRKKSNCV AVCFNDPTTYHEQPSRLKELQDSDIHNFIHLSKNNANORELVNA/KEAKNPDPPNVF LQEKTEENRLSDLKAPEKSGSKI VAGDDPPVGRTKYVNNKKNKONISIKEDPEDNQK SNEESKPNKPSKKRPTANGSGNSGNKKVKLDYRRRVEISQIFRRRTORNLIQ RETPTTEIKETHELLNRIYENSOTKRPFPDLKALRESKHLKLLKAI VNDPDLGSEFH PLCKEILSWADLITELKKELQALPTP"	
CDS	11123..11737 /gene="YLR456W" 11123..11737 /gene="YLR456W" /codon_start=1 /evidence=not experimental /product="Yl456wp" /protein_id="AAB64720.1" /db_xref="GI:717073"	
gene	/translation="MKLNQIIPKDLLRLKSKYKVVHATCSSNCIPSVSLMHYIPVSS ASTFKHYSIIBDCNDYIIFTVFKSVTFRNMSNPVALLFHDWITAKNLTLRKKS VSKDHFSESESTKFNFLDLNQLSELNOVSATINGIADI VNPNSSESTYVRRLL TVNPDAIDILGEDTALIKVNIQKIKVSDMENNSTYGTQVQPV"	
CDS	complement (11865..12824) /gene="NBP1" complement (11865..12824) /gene="NBP1" /note="YLR457C; Weak similarity in central portion of the protein to tropomyosin alpha chain (Swiss Prot. accession number P18442) and kinesin heavy chain (Swiss Prot. accession number P21613)" /codon_start=1 /evidence=not experimental /product="Nbp1p" /protein_id="AAB64721.1" /db_xref="GI:717076"	
gene	/translation="MLKSVQGLWQKQPFGRDGRKREYSGLSDEVKRSALRSRRKQMR PTGSKVLKPRKVTDRTEKIRTNRRKTPKRLTKI POTIRDFVSDNENMSKQNV CDGMTRILKRSQGRPSYNDTTAKSRLRSAPKRISELKYNKQRISELRSQSDG SSGQRNLKYLDREILLQRIKQDEKIKALSKLSQELALYNKRIKRIELLDLD SNHIPSYYTKSRSTMSNLARENDEIKPLKIDLSPSPIRTNSLTSSPMKTYNRDGN PMQPIQENISPACTPPYRSRETEKETLSPISVDF8SYLS"	
CDS	13387..14571 /gene="CDC91" /db_xref="SGD:L0000290" 13387..14571 /gene="CDC91" /notes="YLR459W" /codon_start=1 /evidence=not experimental /product="Cdc91p" /protein_id="AAB64722.1" /db_xref="GI:717072"	
gene	/translation="MDSTALKVALGCIAIRLVANSFLPSLQOQDQSVFSTPTVTSFR SLQEGYLLRNINQVYHGVHHPPIIFELSPNSDRILISILYALIDGLIAYQVTEV TKAFNKLKUKWLPGLLYAVNPVLLTLCISRSIIITFNFAISSSLYCIAGNVLLSS VMISGYLSVYPIPLLIPGLWKSQRORISAIVSILSLILLLLFSYSLGSSQWS FLTQVYGSITTEKFPNGLWMYFPIEMFPIPEFKAFNIFIAVFIPTPTLRVHK QPFYAFILCIGWILVTKPYPSLGDAGFFPFPFPLFGYLRYPYIISALLFHLAVL APIFYHLWVLGSGNSNFFYISLVYALIASILVDLNWMLRIEYDNGIPNPKLVK QI"	
gene	complement (15810..16940) /gene="YLR460C" complement (15810..16940) /gene="YLR460C"	



Query Match 32.3%; Score 359; DB 8; Length 29634;  
Best Local Similarity 58.9%; Pred. No. 9.1e-70;  
Matches 658; Conservative 0; Mismatches 450; Indels 9; Gaps 2;

QY 5 CAGTTCACACCACTCAAAAGCCGCTCATATTGAAGGTGACAAAGCTGTTGTTAAACAG 64  
DB 16930 CAATTCCAGAACCACTGAAGGCTGCTGTCATTGAAGACGGTAAAGCGGTTGTTAAAGAG 16871

QY 65 ATGCTCAGTTCACAGATTAAAGGAGGTACAGCTTGGTGAAGTTGAGCGCTGTCGT 124  
DB 16870 GCATTCCCATCTCGAATTGGAAGAGGATTGCTGATTGAATTAAGACACTCGCTGTCGT 16811

QY 125 GTAACCAACATGATTGGAAGCATATTGCTTATAGATTGGTCCAGAGGTTCAATTCTAG 184  
DB 16810 GTAACCAACATGATTGGGCACACATTGACTACAGATCGGGCTCAAGGATCTATTCTG 16751

QY 185 GATGTGACATGCTGGTACAGTTGTCAAACTGTGACCAAAATGCTAGTACT---GACTGA 241  
DB 16750 GATGTGATGCTGCTGGCCAAATTTGCAAAATGGGCCAGCTGCTCAATCTTAAAGACTTTT 16691

QY 242 AGTTGGAGATACCGGTTTCGGTTTCTTCAAGTGTCTCCCAACAGATCCTTAAAGATG 301  
DB 16630 GTGCTTTTGTGATATCTGCTACTATTTCAACTGTGGTGTGCTACAAATCACCACATGAAC 16571

QY 356 TAACCTACTCAACTGCTGATGAATTTCTGAAGCCCTGTGAAGAACTTCGAATCTGCTG 415  
DB 16570 TCAAAATTTTGGGTGAGGATGTTCTACCTGCGGCCCTGTGAGTCTTTGGAAGGTGAG 16511

QY 416 CATCATTTGCCAGTTTCTGTTGACACTGCTGGTGTAGTTGTGTCATCATCTTGGCTCAA 475  
DB 16510 CCACTATCCCAGTGTCTGACCAACGCGCTTGGTGTGACCTATAACTTGGGCTGG 16451

QY 476 AATGGAATGGCACCCTATCTACCCGCAACATACTCATCATTTATGATTGTTGGGTGGTG 535  
DB 16450 ACCTGAAGTGGAGCCATCAACCCCAAGAAAGCCCATCTTATATGTTGGGCGGTG 16391

QY 536 CTACAGCAGTGGGTCACAACTAATCAAGTGTGCCAAACATCAATGCTTATCTAAGA 595  
DB 16390 CAACGTGAGTGTGAGTGTGCTGCTATCCAAATAGCCAAATAAATGAAATGCTTCAACAGA 16331

QY 596 TTGTAACTGTGCTCTCAAAAGCATGAAAGCTTTTAAAGTCTTATGGTCTGATGATG 655  
DB 16330 TCAATTTGTGGCTTCTCGGAAGCAGCAAGAAACCTTTTGAAGAAATATGGTGTGATGAAT 16271

QY 656 TCTTTGACTATCATGATGACAGGCGTTATTGAGCAGATCAAAATCGAAGTATCCAAACCTGC 715  
DB 16270 TATTTGATTATCATGATTTGACGTGTGTAGACAAATTAACACAGTACACATATCT 16211

QY 716 AACATGTTATGACGCTGTGGGAAGCAGATAGTATCCCGAGGCTTATAAAGTCACAG 775  
DB 16210 CGTATTTAGTCGACTGTGTCGGAATCAAGATACGCTTCAACAGTGTACAAATGTGCG 16151

QY 776 CAGATGCTACCTGCCACATTTATGAAGTGTGTTCCATGACCATTTGAAGCATTCCTG 835  
DB 16150 CCGATAACAGGATGCTACAAATTTGTAATTTAAATAATTTGACAGAGAAACGTCACAAA 16091

QY 836 AAGAAATCAGAAAGATAATGTTTAAATTCATATTCTTTGTTGATCGGTCATCTGGTC 895  
DB 16090 AAGAAACAGAGACAAACGTTACTATTGACATATAAGGCTATTATCAATAGGTGGCC 16031

QY 896 AAGAAATCTTATTTGGGTGCAACAGATTTCTCTGCTAGTCCAGAAATATCATGAAGCCACAG 955  
DB 16030 ATGAAGTACCATTGTGGAACATTTACTTTTACAGCGACTCAGAAGCTAGGAAAGCTGCAA 15971

QY 956 TTAATTCGTTAGTTTATAAATCCACACCTTTAACACGTTGATATCCATCATATGAATA 1015  
DB 15970 TAAATTTTATCAATTCATCAATCCAAAGATTAATGATGACAAATTCGCCATATTTCCAG 15911

QY 1016 TTAAGTTTTTCAGCAACCGCTTAGATGATGTCCAGCTCTCCAGTGAAGGTATATAAAGAG 1075  
DB 15910 TAAGGCTCTATRAAGAACGGCTTTGTGATGTTCTCTCATATCTCTAAAGACATCAATATG 15851

QY 1076 GTAAACCAAAATGTTAAGTATGTTGCCAGGTTATA 1112  
DB 15850 GTAAGAACTCTGCTGAAAAAACTCGTTGCGGTATATAA 15814

RESULT 8  
AX524813 AX524813 1145 bp DNA linear PAT 21-NOV-2002  
LOCUS Sequence 3 from Patent EP1236796.  
DEFINITION  
ACCESSION AX524813  
VERSION  
KEYWORDS AX524813.1 GI:25169909  
SOURCE Saccharomyces cerevisiae (baker's yeast)  
ORGANISM Saccharomyces cerevisiae  
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

REFERENCE 1  
AUTHORS Yamamoto, H. and Kimoto, N.  
TITLE Novel enzyme reductases isolated from kluyveromyces lactis, methods for producing same, and methods for selectively reducing a carbon-carbon double bond of an alpha, beta-unsaturated ketone using the reductases  
JOURNAL Patent: EP 1236796-A 3 04-SEP-2002;  
Daicel Chemical Industries, Ltd. (JP)

FEATURES  
Location/Qualifiers  
1..1145  
/organism="Saccharomyces cerevisiae"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:4932"  
6..1136  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAD57034.1"  
/db\_xref="GI:25169910"  
/translation="MSASIPETMKAVVUENGKAVUKODIPIPELEGFVLIKTVAVAG  
NPTDWHIDFKIPQGLGCDAGQIVKIGNVDAARFAIDGYIYGVHIGASVRPFS  
NGAFYSAISSETAYKPAEPRLCGKDLPEGVKSLGAVSLVPSLUTAGMILTHS  
FGLDWTKPSKAQRDQPIFLFWGATAVGOMLIQAKLNGFSKIIIVASRRHEKLLKE  
YGAELDFYHDADVIQIKKYNNTPIYVDCVSNFTETQQVYKCAADLDATVVLTV  
LTKEDIBEDRONYSIEGTLLYLIGNDVPFGTFLPADPEYKEAAIKFIKPINPKI  
NDGEIHUPVKYVKNGLDDIPOLLDDIKHRSNGEKLVAVLK"

BASE COUNT 343 a 245 c 256 g 301 t

ORIGIN

Query Match 32.1%; Score 357.8; DB 6; Length 1145;  
Best Local Similarity 58.8%; Pred. No. 2e-69;  
Matches 658; Conservative 0; Mismatches 452; Indels 9; Gaps 2;

QY 4 TCAGTTCCAAACCACTCAAAAGCCGCTCATATTGAAGGTGACAAAGCTGTTGTTAAACA 63  
DB 15 TCGATTCCAGAAACCACTGAAGCCGTTGTCATTGAAATGGCAAGCTGTAGTCAAAACAG 74

QY 64 GATGCTCAGTTCAGATTTAAAGGGGTACAGCTTGGTGAAGTTGAGCTGTGCT 123  
DB 75 GACATTCCTCAATTCCTGAATTTAGAAAGGATTTGTTCTAAATTAAGACTGTCGCCGTTGCC 134

QY 124 GGTAAACCAACATGATTGGAAGCATATTGCTTATAAGATTGGTCCAGAGGTTCAATTTCTA 183  
DB 135 GGTAAACCTTACGATTGGAAACATATTGATTTCAAGATTGGTCTCAAGTGGCCCTCTTA 194

QY 184 GGATGTGACATTTGCTGGTACAGTTGTCAAACTTGGAACCAAAATGCTAGTACTG---ACTTG 240  
DB 195 GCGTGTGATGACGCGCGCCAAATCGTAAAGTTGGGCCCAAAATGTTGATGCTGCAAGCTTT 254

QY 241 AAGGTTGGAGTACCGGTTTCGGTTTGTTCACGTTGCTTCCCAACAGATCCTTAAAT 300  
DB 255 GCCATTTGGTGTATCATTTATGGGGTTATTACGGTGTCTTCAGTGAGGTTCCCTCAAC 314

QY 301 GGTGCATTTGCTGAATATGCCAGGGTTTATCCACCTTTGTTTATACAA-----GAGTAAC 354



||||| 315 GGTGCTTGTCTGAGTACTGCTGCAATTCATCCGAGACTCTTATATAACCCAGCAGAG 374  
QY ||||| 355 TTAACCTCACTCACTGCTGATGAAATTTCTGAAGGCCCTGTGAAGAACTTCGAAATCTGCT 414  
Db ||||| 375 TTTAGATTGTCGGTAAAGCAAGCTACCAAGAGGCCCTGTAATCTTTAGAGGGGCA 434  
QY 415 GCATCAATGCCAGTTTCTGTTGACAACCTGCTGTTAGTTTGTGTCATCACTTTGGGCTCA 474  
Db 435 GTATCCCTCCAGTCTCATTTGACCAAGCTGGTATGATCCTTACACATAGTTTGGCTTG 494  
QY 475 AAATGGAATGGACCACTCTACCCGCAACATCTCATCCATATTTGATTTGGGTGGT 534  
Db 495 GACATGACATGGAAGCCCTCAAGCGCAAGAGATCAACCCATCTTATTTGGGTGGT 554  
QY 535 GCTACAGCAGTGGTCAACAACTTAATCCAGTTGCGCAACATATCAATGCTTATCTAAG 594  
Db 555 GCACCTGCTGTCGACAGTCTTATTCATTTGCGCAAAAACATTAACGGTTTCAGCAAG 614  
QY 595 ATTGTAATGTTCTCTTAAAAAGCATGAAAGCTTTTAAAGTCTTTATGCTGCTGATGAT 654  
Db 615 ATCATCGTCTGCTCTCGTAAACATGAAATAATTTGTTGAAGAGTACGGTGCAGATGA 674  
QY 655 GTCTTTGACATCATGATGACAGCGTTTATGAGCAGATCAATCGAAGTATCCAACTG 714  
Db 675 CTTTTTGACTACCAAGTCTGACGTTATCGAAGATATAAAAGAAAGTACAACACATT 734  
QY 715 CAACATGTTATTACGCTGTCGGAAGCAAGATAGTATCCCGAGGCCCTATAAGTCACA 774  
Db 735 CTTTACTGTTGAGTCTGCTCAACACAGAACTATTCACAGGTGTCAATATGCTGCC 794  
QY 775 GCAGATAGTCTACCTGCCACATATTATAGAAAGTGGTTCATGACCAATTTGAAGCATTCCT 834  
Db 795 GCTGATGACTTAGACGCTACGGCTGTTCAATTGACCGTTTAAACGAAAGATATCAAG 854  
QY 835 GAAGAAATCAGAAAGATAATGTTAAATGATATTACTTTGTTGATTCGTGATCTGCT 894  
Db 855 GAGGAAGACAGGAGGCAAAACGTCAGTATTGAAAGAAACCTCTTATATTGATAGGAGT 914  
QY 895 CAAGAAATTTCTATTGGTGCACAAAGATTTCTGCTAGTCCAGAAATATCATGAAGCCCA 954  
Db 915 AACGAGTCCATTTGCGACGTTTACTTTGCGAGCAGACCTCGAATACAAAGGAGCGCC 974  
QY 955 GTTAAATTCGTTAAATTTAAATCCACACCTTAAACAAAGCTGATATCCATCATATGAAT 1014  
Db 975 ATAAATTTTAAATTATCATATCCAAATCAATGATGCTGTAATCCACCATCCCA 1034  
QY 1015 ATTAAAGTTTTCAGCAAGCGCTTAGATGATGTCCTCAGCTCTCTACTGAAGGTATAAAGAA 1074  
Db 1035 GTGAAAGTTTACAAAGACGCGTTAGATGATATCCACAGTTACTTGTATGATATTAAGCAC 1094  
QY 1075 GGTAAACCAACAAATGTTAAGTATGTTGCCAGGTTATAA 1113  
Db 1095 GGGAGGAATTTCTGGCAAAAGTTGGTGGCCGCTTGAA 1133

## RESULT 9

BD174422 1145 bp DNA linear PAT 18-MAR-2003  
LOCUS Novel enone reductase, process for producing the same, and method  
DEFINITION for selectively reducing carbon-carbon double bonds of alpha,  
beta-unsaturated ketone using the same.

BD174422

BD174422.1 GI:29120110

JP 2002247987-A/2.

Saccharomyces cerevisiae (baker's yeast)

Saccharomyces cerevisiae

Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

Saccharomycetales; Saccharomycetaceae; Saccharomyces.

1 (bases 1 to 1145)

Yamamoto, H. and Kimoto, K.

Novel enone reductase, process for producing the same, and method

for selectively reducing carbon-carbon double bonds of alpha,

JOURNAL beta-unsaturated ketone using the same  
Patent: JP 2002247987-A 2 03-SEP-2002;  
DAICEL CHEMICAL INDUSTRIES LTD  
COMMENT OS Saccharomyces cerevisiae (yeast)  
PN JP 2002247987-A/2  
PD 03-SEP-2002  
PI 23-FEB-2001 JP 2001049363  
PT HIROAKI YAMAMOTO, KUNIHITO KIMOTO  
PC C12N15/09, C12N1/15, C12N1/19, C12N1/21, C12N5/10, C12N9/02, C12P7/  
PC 26//  
PC (C12N15/09, C12R1:645), C12N15/00, C12N5/00, (C12N15/00, C12R1:645)  
CC Novel enone reductase, process for producing the same, and CC  
method for  
CC selectively reducing carbon-carbon double  
bonds of alpha, beta-  
CC ketone using the same  
CC ununsaturated  
FH Key Location/Qualifiers  
FT CDS (6)..(1136).

## FEATURES

source

1..1145

/organism="Saccharomyces cerevisiae"

/mol\_type="genomic DNA"

/db\_xrefs="taxon:4932"

BASE COUNT 343 a 245 c 256 g 301 t

## ORIGIN

Query Match 32.1%; Score 357.8; DB 6; Length 1145;  
Best Local Similarity 58.8%; Pred. No. 2e-69;  
Matches 658; Conservative 0; Mismatches 452; Indels 9; Gaps 2;  
QY 4 TCAGTTCCCAACCACTCAAAAGCCGTCATCATTTGAAGGTGACAAAGCTGTGTAAACA 63  
Db 15 TCAGTTCCAGAAACCATGAAGCCGTTGTCATTTGAAATGGCAAGGCTGTAGTCMAACAG 74  
QY 64 GATGCTCTAGTTCCAGAAATTAAGAGGGGTACAGCCTTGGTGAAGGTTGAGGCTGTGCT 123  
Db 75 GACATTTCAATTCCTGAATTTAGAAGAGGATTTGTTCTAATTAAGACTGTGCGCGTTGCC 134  
QY 124 GGTAAACCACTGATTTGGAAGCATATGCTTATAGATTTGGTCCAGAGGTTCAATCTA 183  
Db 135 GGTAAACCTTACCGATTTGGAAGCATATGATTTCAAGATTTGGTCTCAAGGTCCTCTTA 194  
QY 184 GGATGTGACATTTGCTGGTACAGTTGTCAAACTTGGACCAAACTGTAGTACTG---ACTTG 240  
Db 195 GGCTGTGATGACAGCGCGCCAAATCGTAAAGTTGGGCCAAATGTTGATGCTGCACGCTTT 254  
QY 241 AAGTTGGAAGATACCGGTTTTCGGTTTTCACGGTGTCTCCCAACAGATCTCAAAAT 300  
Db 255 GCCATTTGGTGTATACATTTATGGGTTATTCACGGTGTCTTCAAGTGTCTTCCCTCAAC 314  
QY 301 GGTGCAATTTGCTGAATATGCCAGGTTTATCCACCTTTGTTTACAA-----GAGTAAC 354  
Db 315 GGTGCTTTGCTGAGTACTCTGCCATTTTCATCCGAGACTGCTTATATAACCCAGCAGAG 374  
QY 355 TTAACCTCACTCAACTGCTGATGAAATTTCTGAAGGCCCTGTGAAGAACTTCGAATCTGCT 414  
Db 375 TTTAGATTGTCGGTAAAGCAAGCTACCAAGAGGCCCGGTAAATCTTTAGAGGGGCA 434  
QY 415 GCATCAATGCCAGTTTGGTGAACACTGCTGTTAGTTGTTGTCATCACTTTGGGCTCA 474  
Db 435 GTATCCCTCCAGTCTCAATTGACCGGCTGGTATGATCCTTACACATAGTTTGGCTTG 494  
QY 475 AAATGGAATGGACCACTCTACCCGCAACATCTCATCCATATTTGATTTGGGTGGT 534  
Db 495 GACATGACATGGAAGCCCTCAAGCGCAAGAGATCAACCCATCTTATTTGGGTGGT 554  
QY 535 GCTACAGCAGTGGTCAACAACTTAATCCAGTTGCGCAACATATCAATGCTTATCTAAG 594  
Db 555 GCACCTGCTGTTGGCCAGATGCTTATTCATTTGCGCAAAAACATTAACGGTTTCAGCAAG 614  
QY 595 ATTGTAATGTTCTTAAAAAGCATGAAAGCTTTTAAAGTCTTTATGCTGCTGATGAT 654

Db 615 ATCATCGTCGTTCTCTCGTAACATGAAAAATTTGTTGAAGAGTACGGTGCAGATGAA 674  
QY 655 GTCTTTGACTATCATGATGAGCGGTTATTGAGCAGATCAAAATCGAAGTATCCAAACCTG 714  
Db 675 CTTTTTGACTACACGATGCTGAGTTATCGAAGATAAAAGAGTACAAACAT 734  
QY 715 CAACATGTTATTGACGCTGTTGGAAGCGAAGATAGTATCCCGAGGCTATAAAGTCACA 774  
Db 735 CTTTACTTGGTGGACTGTCTCCAAACACAGAACTATTCAAACAGGTGTACAAATGTGCC 794  
QY 775 GCAGATAGTCTACCTGCCCAATATTAGAGTGGTCCAAATGACCAATGAAAGCATTCCT 834  
Db 795 GCTGATGACTTAGACGCTACGGTCTGTTCAATTGACCGTTTAAACCGAAAGATATCAAG 854  
QY 835 GAAGAAATTCATGAAAGATATGTTAAATTTGATATTTGTTGTTGATCGTGCATCTGGT 894  
Db 855 GAGAGACAGAGGAGGAAAACGTCAGTATTGAAGAACCCCTTCTATTTGATAGAGGT 914  
QY 895 CAAGAAATTCATTTGGGTGCAACAAGATTTCTGCTAGTCCAGAAATATCATGAAGCCACA 954  
Db 915 AACGAGTCCCAATTTGSCACGTTTACTTTGCCAGCAGACCCCTGAATACAGGAAGCGCC 974  
QY 955 GTTAAATTCGTTAAGTTTAAATTCACACCTTAACAAACGCGTATATCATATATGAAT 1014  
Db 975 ATAAATTTTAAAGTTTCATCAATCCAAATAATCAATGATGTTGAAATCCACCATCCCA 1034  
QY 1015 ATAAAGTTTTCAGCAACGCTTAGATGATGCCAGCTCTCACTGAAGGTATAAAGAA 1074  
Db 1035 GTGAAGTTTACGAAGACGGTTAGATGATATCCACAGTTACTTGATGATATTAAGCAC 1094  
QY 1075 GGTAAGAAACAAAATTTAAAGTATGTTGCCAGGTATATA 1113  
Db 1095 GCGAGGAATTCGCGGAAGATGTTGTCGCTTGAAA 1133

RESULT 10  
AX536730  
LOCUS AX536730 1631 bp DNA linear PAT 22-NOV-2002  
DEFINITION Sequence 331 from Patent WO02064766.  
ACCESSION AX536730  
VERSION AX536730.1 GI:25263166  
KEYWORDS  
SOURCE Saccharomyces cerevisiae (baker's yeast)  
ORGANISM Saccharomyces cerevisiae  
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
Saccharomycetales; Saccharomycetaceae; Saccharomyces.

REFERENCE 1  
Contreras, R.H., Eberhardt, I., Luyten, W.H. and Reekmans, R.J.  
Bax-responsive genes for drug target identification in yeast and  
fungi  
JOURNAL Patent: WO 02064766-A 331 22-AUG-2002;  
JANSSEN PHARMACEUTICA N.V. (BE)  
FEATURES  
Location/Qualifiers  
1.1631  
/organism="Saccharomyces cerevisiae"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:4932"  
BASE COUNT 477 a 357 c 346 g 451 t  
ORIGIN

Query Match 32.1%; Score 357.8; DB 6; Length 1631;  
Best Local Similarity 58.8%; Pred. No. 2e-69;  
Matches 658; Conservative 0; Mismatches 452; Indels 9; Gaps 2;  
QY 4 TCAGTTCCAAACCACTCAAAAGCGGTCATCATTTGAAGGTGACAAAGCTGTTGTTAAACA 63  
Db 510 TCAGTTCCAAACCACTCAAAAGCGGTTGTCATTTGAAATGGCAAGGCTGTAGTCAACAG 569  
QY 64 GATGTCCTCAGTTCCAGAAATTTAAAGGAGGTACAGCCCTTGGTGAAGGTGTAGGCTGTGCT 123  
Db 570 GACATTCCAATTCCTGAATTTAGAGAGGATTTGTTCTAATTAAGACTGTCGCGCTGCC 629  
QY 124 GGTAACCAACTGATTGGAAGCATATTGCTTATAAGATTGGTCCAGAAAGGTTCAATCTA 183

Db 630 GGTAACCCCTACCGATTGGAAACATATTGATTTCAAGATTGGTCTCAAGGTGCGCTCTTA 689  
QY 184 GATGTGACATTTGCTGGTACAGTTGTCAAACTTGGACCAAAATGCTAGTACTG---ACTTG 240  
Db 690 GGCTGTGATGTCAGCGCGGCCAAATCGTAAAGTTGGGCCCAAAATGTTGATGCTGCACGCTTT 749  
QY 241 AAGGTTGGAGATACCGGTTTTCGTTTTCAGGTTCTTCCAAACAGATCTCTAAAAAT 300  
Db 750 GCCATTGGTGATTACATTTATGGGTTATTACGGTGCTTCAGTGAGGTTCCCTCAAC 809  
QY 301 GGTGCATTTGCTCAATATGCCAGGGTTTATCCACCTTTGTTTACAA-----GAGTAAC 354  
Db 810 GGTGCTTTGCTGAGTACTCTGCCATTTTCATCCGAGACTGCTTATAAACCCAGCCAGAGAG 869  
QY 355 TTAACCTCACTCACTGCTGATGAAATTTCTGAAGGCCCTGTGAAGNACTTCAATCTGCT 414  
Db 870 TTTAGATTGTGGGTAAAGACAAGCTTACCAAGGCCCGGTAAATCTTTAGAAAGGGCA 929  
QY 415 GCATCATTTGCCAGTTTTCGTTGACACTGCTGTTGTTAGTTTGTGTCATCACTTTGGGCTCA 474  
Db 930 GTATCCCTCCAGTCTCATTTGACCGCTGTTATGATCCTTACACATAGTTTGGCTTG 989  
QY 475 AAAATGGAATGGCACCCCATCTACCCCGCAACATACTCATCTCCATTTATTTGGGCTGGT 534  
Db 990 GACATGACATGGAAGCCCTCCAAAGCGCAAGAGATCAACCCCATCTTTATTTGGGCTGGT 1049  
QY 535 GCTACAGAGTGGGTCAACAACTAATTCGAAGTTGCCAAACATATCAATGCTTATATAAG 594  
Db 1050 GCCACTGCTGTTGGCCAGATGCTTATTTCAATTTGGCAAAAAAACTAAACGGTTTCAGCAAG 1109  
QY 595 ATTGTAACCTGTTCTTCTAAAAAGCATGAAAGCTTTTAAAGTCTTTATGCTGCTGATGAT 654  
Db 1110 ATCATGCTGTTGCTTCTGTAACATGAAAATTTGTTGAAGAGTACGGTGCAGATGAA 1169  
QY 655 GTCTTTGACTATCATGATGACGGCTTATTGAGCAGATCAAAATCGAAGTATCCAAACCTG 714  
Db 1170 CTTTTTGACTACCGAGTGTGACGTTATCGAAGATGATAAAAGAAAGTACAAACATTT 1229  
QY 715 CAACATGTTATTGACGCTGTGGGAAGCGAAGATAGTATCCCGAGGCTTATAAGTCACA 774  
Db 1230 CCTTACTTGGTGGACTGTGCTCTCAACACAGAAACTATTCAACAGGTGTACAAATGTGCC 1289  
QY 775 GCAGATAGTCTACCTGCCACATTTATTAGAAGTGGTTTCCAATGACCATTCGAAAGCATTCCT 834  
Db 1290 GCTGATGACTTAGACGCTACGGTCTGTTCAATTGACGTTTTTAAACCGAAAGATATCAAG 1349  
QY 835 GAAGAAATCAGAAAAAGATAATGTTAAATTTGATATTTACTTTTGTGTCGTCATCTGGT 894  
Db 1350 GAGGAAGACAGGAGGCAAAACGTCAGTATTGAAGGAACCCCTTCTATATTGATAGGAGT 1409  
QY 895 CAAGAAATTCATTTGGGTGCAACAAGATTTCTGCTAGTCCAGAAATATCATGAAGCCACA 954  
Db 1410 AACGACGTCCTCAATTTGGCAGCTTTTCTTTGGCAGCAGACCCCTGAAATACAAAGAACCGCC 1469  
QY 955 GTTAAATTCGTTAAGTTTATAAATCCACACCTTTAAACAGGGTATCCATCATATGAAT 1014  
Db 1470 ATAAATTTATTAGTTTCAATCAATCCAAATATCAATGATGGTGAATCCACCATCCCA 1529  
QY 1015 ATTAAGTTTTCAGCAACGGCTTAGATGATGCCAGCTCTCACTGAAGGTATATAAAGAA 1074  
Db 1530 GTGAAAGTTTCAAGAAAGCGGTAGATGATATCCACAGTTACTTGTGATGATATTAAGCAC 1589  
QY 1075 GGTAAAAACAAAATGTTAAGTATGTTGCCAGGTATATA 1113  
Db 1590 GGGAGGAATTCCTGGCAAAAAGTTGGTTGCCGCTTTGAAA 1628

RESULT 11  
SCYNL134C/c  
LOCUS SCYNL134C 1944 bp DNA linear PLN 11-AUG-1997  
DEFINITION S.cerevisiae chromosome XIV reading frame ORF YNL134C.  
ACCESSION Z71410 Y13139

```

VERSION 271410.1 GI:1302075
KEYWORDS Saccharomyces cerevisiae (baker's yeast)
SOURCE Saccharomyces cerevisiae
ORGANISM Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE 1 (bases 1 to 1944)
AUTHORS Mallet, L., Bussereau, F. and Jacquet, M.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1944)
AUTHORS MIPS.
TITLE Direct Submission
JOURNAL Submitted (29-APR-1996) Data collected by MIPS on behalf of the
European yeast chromosome XIV sequencing project. MIPS at the
Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a D-82152
Martinsried, FRG; E-mail: Mewes@mips.embnat.org
FEATURES
    Location/Qualifiers
        1..1944
            /organism="Saccharomyces cerevisiae"
            /mol_type="genomic DNA"
            /db_xref="taxon:4932"
            /chromosome="XIV"
            complement(225..1355)
            /note="ORF YNL134c"
            /codon_start=1
            /protein_id="CAA96016.1"
            /db_xref="GI:1302076"
            /db_xref="SWISS-PROT:P53912"
            /translation="MSASIPETMKAVIENGKAVVQKDIPIPELEGVLIKTVAVAG
            NPTDKRIDFKIGQALGCDAGQIVKLGPNVDAARFAIGDTIYGVHIGASVRPFS
            NGPAEYSAISSETAYKPAERFRCGLKPEGVKSLGALVSLPVTAGMILTHS
            FGLDWTWPKAQORDQPILEWGGATATVGOMLIQIAKLGFSKIIIVASRHEKLLKE
            YGADLPDYHDADVIEQIKKKYNNIPYLVCVSNTEIIQOVYKCAADLDATVVLTV
            LTKEDIKEEDRRONVSIEGTLVYLGNDVPFGFTLTPADPEYKEAKIPIKINPKI
            NDGEIHIPVKYNGKLDIPQLDDDIKHGNSGEKLVAVLK"
BASE COUNT 542 a 397 c 420 g 585 t
ORIGIN
Query Match 32.1%; Score 357.8; DB 8; Length 1944;
Best Local Similarity 58.8%; Pred. No. 2e-69;
Matches 658; Conservative 0; Mismatches 452; Indels 9; Gaps 2;
Qy 4 TCAGTTCAACCACTCAAAAGCGGTGATCATCTGAAGGTGACAAAGCTGTGTGTTAAACA 63
Db 1346 TCGATTCCAGAAACCATGAAAGCGGTGTGTCATTGAAATGCGAAGGCTGTAGTCAACAG 1287
Qy 64 GATGCTCAGTTCAGAAATTAAGAGGAGGTACACCTTGGTGAAGGTTGAGGCTGTGCT 123
Db 1286 GACATTCGAATTCCTGAATAGAGAGGATTTGTTCTTAATTAAGACTGTGCGCGTTGCC 1227
Qy 124 GGTAACCACTGATTCGAAGCATATGCTTATGAAGATTGTCAGAGAGGTTCAATTTCTA 183
Db 1226 GGTAACTCTACGATTCGAACATATGATTTCAAGATTGCTCTCAAGGTGCGCTCTTA 1167
Qy 184 GGATGTGACATTCGTGTGATAGTTGTCAAACTTGACCAAAATGCTAGTACTG---ACTTG 240
Db 1166 GGCTGTGATGAGCGCGCAATGTAAGTTGGGCCCAATGTTGATGCTGACGCTTT 1107
Qy 241 AAGTTGAGATACCGGTTTCGGTTTGTGTCACCGTGTCTCCAAACAGATCTCAAAAT 300
Db 1106 GCCATTGGTATTAATTTATGGGGTTATTACCGGTGCTTCAGTGAGGTTCCCTCAAC 1047
Qy 301 GGTCATTTGCTGAATATGCCAGGTTTATCCACCTTTGTTTACAA-----GAGTAAC 354
Db 1046 GTGCTTTGCTGAGTACTCTGCCATTTATCCAGACTGCTTATAACACGACGAGAG 987
Qy 355 TTAATCACTCACTGCTGATGAATTTCTGAAGGCCCTGTGAAGAACTTCGAATCTGCT 414
Db 986 TTTAGATTGTCGGTAAAGACAAGCTACCAAGAGCGCCCGTAAATCTTTAGAGGGCA 927
Qy 415 GCATCATTCGAGTTTGTGACAACTGCTGGTGTAGTTTGTGTGATCACTTGGGCTCA 474
Db 926 GTATCCCTCCAGTCTCATTTGACCACGCGTGTATGATCTCTTACACATAGTTTGGCTTG 867

```

---

```

475 AAATGGAATGGCACCCTATCTCCCGCAACATATCTCATCTCATCTATTTGTTGGTGGT 534
866 GACATGACATGGAAGCCCTCCAAAGCGCAAGAGATCAACCCATCTTATTTGGGTGGT 807
535 GCTACAGCAGTGGGTCAACAACTAATCCAAAGTTGCCAAACATATCAATGCTTATTAAG 594
806 GCCACTGCTGTGGCCAGATGCTTATCAATTGGCAAAAATAAACCGTTTCAGCAAG 747
595 ATTGTAACGTTCTCTTAAAGAGCATGAAAGCTTTTAAAGCTTTATGCTGCTGATGAT 654
746 ATCATGCTGCTGCTCTCTGTAACATGAAATTTGTTGAAAGAGTACGGTGCAGATGAA 687
655 GTCTTTGACTATCATGATGACGCGTTTATGACGAGATCAAAATCGAAGTATCCAAACCTG 714
686 CTTTTTGACTACCAAGTCTGAGCTTATCGAACAGATATAAAGAAAGTACAAACATTT 627
715 CAACATGTTATTGACGCTGGGAAGGAGATAGTATCCCGAGGCTTATTAAGTCACA 774
626 CTTTACTTGGTGAAGTGTCTCCAAACACAGAAACTATTCAACAGGTTGTAACAATGTGCC 567
775 GCAGATAGTCTACCTGCCACATTTATGAGTGGTTCCCAATGACCATTTGAAGCATTTCT 834
566 GCTGATGACTTAGCGCTACGCTCGTCAATTCACCGTTTTAACCGAAAGATATCAAG 507
835 GAAGAATCAGAAAGATATGTTAAATTTGATATTACTTTGTTGATCGTCATCTGGT 894
506 GAGGAGACAGGAGGCAAAACGTCAGTATTTGAAGGAACCTTCTATATTTGATAGGAGT 447
895 CAAGAATTTCTATTGGGTGCAACAAAGATTTCTGTAGTCCAGAAATATCATGAAGCACA 954
446 AACGAGCTCCATTTGGCAGTTTACTTTGCCACAGACACCTGAATACAAAGAAAGCGGCC 387
955 GTTAAATTCGTTAAGTTTATATATCCACACTTAACAAAGCGTGATATCCATCATATCAAT 1014
386 ATAAATTTATTAAAGTTTCATCAATCCAAATCAATGATGGTGAATTCACCATATCCCA 327
1015 ATTAAAGTTTTCAGCAACGCTTAGATGATGTCCTCCAGCTCTCACTGAAGGTATATAAGAA 1074
326 GTCAAGTTTACAAAGACGGTTAGATGATATCCACAGTTACTTGTGATGATTTAAGCAC 267
1075 GGTAAACCAAAAATGTTAAGTATGTTGCCAGGTTATAA 1113
266 GCGAGGAATTTCTGCGAAAAGTTGGTTGCGCTTGA 228

```

---

```

RESULT 12
SCCHXIV43/c
LOCUS S.cerevisiae chromosome XIV DNA (43.5 kb). linear PLN 17-FEB-1997
DEFINITION
ACCESSION 246843
VERSION 246843.1 GI:861113
KEYWORDS adenosine deaminase; cyclase-associated protein; FR506-binding
protein proline rotamase; mating pheromone alpha-factor 2;
mitochondrial outer membrane component; mitochondrial ribosomal
protein; NH3 permease; Phe-tRNA; sn-1,2-diacylglycerol
cholinephosphotransferase; tyrosine phosphatase.
SOURCE Saccharomyces cerevisiae (baker's yeast)
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE 1 (bases 25310 to 25400)
AUTHORS Bull, P., Thorikay, M., Moenne, A., Wilkens, M., Sanchez, H.,
Valenzuela, P. and Venegas, A.
TITLE The yeast tRNA (Phe) gene family: structures and transcriptional
activities reveal member differences not explained by intragenic
promoters
JOURNAL DNA 6 (4), 353-362 (1987)
MEDLINE 88003973
PUBMED 3308382
REFERENCE 2 (bases 2855 to 2968)
AUTHORS Michaelis, S. and Herkowitz, I.
TITLE The a-factor pheromone of Saccharomyces cerevisiae is essential for

```

JOURNAL MEDLINE PUBMED	Mol. Cell. Biol. 8 (3), 1309-1318 (1988)	mating	Jacquet,M. and Pfanner,N.
REFERENCE	88216607		The mitochondrial receptor complex; Mom22 is essential for cell
AUTHORS	3285180		viability and directly interacts with preproteins
TITLE	3 (bases 30002 to 31314)		Mol. Cell. Biol. 15 (1995) in press
	Hjeltnad,R.H. and Bell,R.M.		12 (bases 7894 to 9390)
	The sn-1,2-diacylglycerol cholinephosphotransferase of		Marini,A. and Andre,B.
	Saccharomyces cerevisiae. Nucleotide sequence, transcriptional		MEP2
	mapping, and gene product analysis of the CPT1 gene		Unpublished
JOURNAL MEDLINE PUBMED	J. Biol. Chem. 265 (3), 1755-1764 (1990)		13 (bases 1 to 43481)
REFERENCE	90110247		Direct Submission
AUTHORS	2153142		Jacquet,M.
TITLE	4 (bases 17182 to 18759)		Submitted (29-NOV-1994) Michel Jacquet, Institut de Genetique et
	Fedor-Chaiken,M., Deschenes,R.J. and Broach,J.R.		Microbiologie, Universite, Paris-aud, Laboratoire Information
	SRV2, a gene required for RAS activation of adenylate cyclase in		Genetique et developpement, Orsay, 91405, FRANCE
	yeast		On Jun 14, 1995 this sequence version replaced gi:854489.
JOURNAL MEDLINE PUBMED	Cell 61 (2), 329-340 (1990)		Location/Qualifiers
REFERENCE	90235282		1. 43481
AUTHORS	2158860		/organism="Saccharomyces cerevisiae"
	Field,J., Vojtek,A., Ballester,R., Bolger,G., Colicelli,J.,		/mol_type="genomic DNA"
	Ferguson,K., Gerst,J., Katoka,T., Michael,T., Powers,S.,		/strain="alpha S288C"
	Riggs,M., Rodgers,L., Wieland,I., Wheland,B. and Wigler,M.		/db_xref="taxon:4932"
TITLE	Cloning and characterization of CAP, the S. cerevisiae gene		/chromosome="XIV (43.5kb fragment)"
	encoding the 70 kd adenylyl cyclase-associated protein		/clone="alpha 14-12"
JOURNAL MEDLINE PUBMED	Cell 61 (2), 319-327 (1990)		/clone_lib="from Peter Philippsen"
REFERENCE	90235281		complement (348. .1112)
AUTHORS	2184942		/citation=[10]
	6 (bases 23226 to 22667)		/codon_start=1
	Wiederrecht,G., Brizuela,L., Elliston,K., Sigal,N.H. and		/label=JTD254
	Siekterka,J.J.		/product="cytoskeleton-associated protein (putative)"
TITLE	FKB1 encodes a nonessential FK 506-binding protein in Saccharomyces		/protein_id="CAA86878.1"
	cerevisiae and contains regions suggesting homology to the		/db_xref="GI:854490"
	cyclophilins		/translation="WVRVIESELVTEKELNSLKIROFKDLRYHVTGVEDMEIV
JOURNAL MEDLINE PUBMED	Proc. Natl. Acad. Sci. U.S.A. 88 (3), 1029-1033 (1991)		VRQYDNKEIYTKGGAYSNEDANFLKGEELIVVTDNSAGISNQLATQAEGI
REFERENCE	91126049		PSMEVISEDYLRDQSVLRWVAHGYGRFNAQQSRAALAKQDEAYAREQLTAAIG
AUTHORS	1704127		RHCRTVDGSAPEAILRYVGLPLDVMGTGCVFPEAAGKNGCRINGVTLFGVPAP
TITLE	7 (bases 22326 to 22667)		GHGSFVRPAVEILSKDEESAEDVDHDDVESDDEI"
	Heitman,J., Movva,N.R., Hiestand,P.C. and Hall,M.N.		join (1381. .1398, 1519. .1824)
	FK 506-binding protein proline rotamase is a target for the		/gene="orf2"
	immunosuppressive agent FK 506 in Saccharomyces cerevisiae		join (1381. .1398, 1519. .1824)
JOURNAL MEDLINE PUBMED	Proc. Natl. Acad. Sci. U.S.A. 88 (5), 1948-1952 (1991)		/gene="orf2"
REFERENCE	91156723		/note="putative member of the Sm protein family"
AUTHORS	1705713		/citation=[10]
	8 (bases 22326 to 22667)		/codon_start=1
	Koltin,Y., Faucette,L., Bergema,D.J., Levy,M.A., Cafferkey,R.,		/label=JTA107
	Koser,P.L., Johnson,R.K. and Livi,G.P.		/protein_id="CAA86879.1"
TITLE	Rapamycin sensitivity in Saccharomyces cerevisiae		/db_xref="GI:854491"
	peptidyl-prolyl cis-trans isomerase related to human FK506-binding		/db_xref="SWISS-PROT:P53905"
	protein		/translation="WHQHSQRKKEGPKREAILDLAKYKDSKIRVKLMGGKLVIGVL
JOURNAL MEDLINE PUBMED	Mol. Cell. Biol. 11 (3), 1718-1723 (1991)		KGYDLMNLVLDDTVEYMSNPDENNELISKNARKIGLTVIRGTLVLSLSAEGSDV
REFERENCE	91141524		LYMQK"
AUTHORS	1796117		1399. .1518
	9 (bases 19039 to 20496)		/gene="orf2"
	Boguta,M., Dmochowska,A., Borsuk,P., Wrobel,K., Gargouri,A.,		2156. .2458
	Lazowska,J., Slonimski,P.P., Szczesniak,B. and Kruszevska,A.		/gene="orf3"
TITLE	NAM9 nuclear suppressor of mitochondrial ochre mutations in		2156. .2458
	Saccharomyces cerevisiae codes for a protein homologous to S4		/gene="orf3"
	ribosomal proteins from chloroplasts, bacteria, and eucaryotes		/citation=[10]
JOURNAL MEDLINE PUBMED	Mol. Cell. Biol. 12 (1), 402-412 (1992)		/codon_start=1
REFERENCE	92107194		/label=JTB100
AUTHORS	1729612		/protein_id="CAA86880.1"
	10 (bases 1 to 43481)		/db_xref="GI:854492"
	Mallet,L., Busseureau,F. and Jacquet,M.		/db_xref="SWISS-PROT:P53906"
	A 43.5 kb segment of yeast chromosome XIV, which contains MFA2,		/translation="MSNTKHTSHMELIRIILRLFLIMLIFRNSVSKWTFOEL
	MEP2, CAP/SRV2, NAM9, FKB1/FPRI/RBP1, MOM22 and CPT1, predicts an		LPFRYKNSNSVNNRPFSENLVDVDDVMYDKTRLFILFPFSTITTFMV"
	adenosine deaminase gene and 14 new open reading frames		2855. .2971
JOURNAL MEDLINE PUBMED	Yeast 11 (12), 1195-1209 (1995)		/gene="MFA2"
REFERENCE	86109932		2855. .2971
AUTHORS	8619318		/gene="MFA2"
	11 (bases 29208 to 29663)		/function="induces mating processes on alpha responsive
	Honlinger,A., Kuebrich,M., Moczko,M., Gaertner,F., Mallet,L.,		cells"
	Busseureau,F., Eckerskorn,C., Lottspeich,F., Dietmeier,K.,		/citation=[2]
			/citation=[10]

```
/codon_start=1
/label=JTB38
/product="mating pheromone a-factor 2"
/protein_id="CAA6681.1"
/db_xref="GI:854493"
/db_xref="SWISS-PROT:P34166"
/translation="MQPITASTQATQKDSSEKKDNIYIKGLFWDPAVCVIA"
/complement(3261..5483)
/gene="orf5"
/complement(3261..5483)
/gene="orf5"
/citation=[10]
/codon_start=1
/label=JTD740
/protein_id="CAA6682.1"
/db_xref="GI:854494"
/db_xref="SWISS-PROT:P51907"
/translation="MSSSIFEMIDHGMSTGLKLSQTASSISMGDEFILCSSTNS
ILUSPLKVFNIHIDSITDINTINIMNEIVPQGVDDVDADKNVLYCIDPYPVEPPCY
DFANPSKVIKPIYIEHRCPLTSVKPESYTPSVSEHYTVVSMKMKLYSPFNASRLWN
NFILQINSTQINPYSIDSLRHKIKYRGDMDFDHSKTAGDRHHSARSLNAPT
KSTVQPKYDKERICGIADEKFLSDEKFLKYSYLOCAKVLPIIDYSRDPVLMR
CEGOPLVQSHVDELIVMAYLNMGLSLSLDLREMPVTVSVPBRHPRAEFKOH
NKXNKNKSNQSDRNDRSRSHLLLRSHSTSVVVKVATGNERTPNKRSRSLSLPP
SVSDASHDNGPNDGSLAGESSQTFGLFTSKLRNFFKTDSSSKNSMDIGQKMR
SKELNSVQBEIDNGSTNTNTSVLSFTSPHSLTAQTSIHENFRSRNSNPIDP
LRCDRSVLKNNFEPVYEGTGRSTISNSAERTNREKLNKRGDNVDGDEDEDDA
NYEEDBDGVVEDDDINRLVLDERSRYENADVEYGTIPSTKSFVYTGCLNDFPKK
LSLCKGNKNSNDSKWPATQLVSRKRYIKDSLRCIKPLTEHPWVGKLIIFKRALPP
```

Query Match 32.1%; Score 357.8; DB 8; Length 43481;

Best Local Similarity 58.8%; Pred. No. 1.7e-69;

Matches 658; Conservative 0; Mismatches 452; Indels 9; Gaps 2;

```
Qy 4 TCAGTTCACACCATCAAAAGCCGTCATCTATTCAGAGTGACAAAGCTGTGTTAAACA 63
Db 24013 TCGATTCCAGAAACCATGAAAGCCGTTGTCATTGAAAATGCGAAGCGTGTAGTCAACAG 23954

Qy 64 GATGCTCAGTTCAGAAATTAAGAGGGTACAGCGCTTGGTGAAGGTTGAGCGCTGTGCT 123
Db 23953 GACATTCGATCTGATAGAGAGAGGATTTGTTCTAATTAGACTGTGCGCGTTGCC 23894

Qy 124 GGTAAACCAACTGATGGAAGCATATTGCTTATAGATTGGTTCAGAGGTTCAATTTCTA 183
Db 23893 GGTAACCTACCGATTGGAACATATTGATTCAAGATTGGTCTCTCAAGGTGCCCTCTTA 23834

Qy 184 GGATGTGACATTCGTGTGACAGTGTGCAACTTGACCAATGCTAGTACTG----ACTTG 240
Db 23833 GGCTGTGATGACGCGCGCCCAATCGTAAAGTTGGGCCCCAAATGTTGATGCTGCACGCTTT 23774

Qy 241 AAGGTTGAGATACCGGTTTCGGTTTGTTCACGCTGCTCCCAACAGATCCTTAAAAAT 300
Db 23773 GCCATTGGTATACATTTATGGGGTTATTACCGGCTGCTCAGTGAGTTCCCTCAAC 23714

Qy 301 GGTGCATTTGCTGAATATGCCAGGTTTATCCACCTTTGTTTACAA-----GAGTAAC 354
Db 23713 GGTGCTTTGCTGAGTACTCTGCCATTTCATCCGAGACTGCTTATAACACGAGCAGAGAG 23654

Qy 355 TTAACCTACTCACTGCTGATGAAATTTCTGAAGCCCTGTGAAGACTTCGAAATCTGCT 414
Db 23653 TTTAGATTGTGCGTGAAGACAAGCTACCAAGAGCCCGCCGTAATCTTTAGAGAGGGCA 23594

Qy 415 GCATCATTTGCGAGTTTCGTTGACAACTGCTCGTGTAGTTGTTGTCATCATCTGGGCTCA 474
Db 23593 GTATCCCTCCAGTCTCATGACCACGCTGGTATGATCCTTACACATAGTTTGGCGTTG 23534

Qy 475 AAAATGGAATGGACCCCATCTACCCGCAACATACTCATCCATTATTGATTTGGGGTGGT 534
Db 23533 GACATGACATGGAAGCCCTCCAAAGCGCAAGAGATCAACCCATCTATTTTGGGGTGGT 23474

Qy 535 GCTACAGCAGTGGGTCAACAACTAATCAAGTTGCCAAACATATCAATGCTTATCTAAAG 594
Db 23473 GCCACTGTGTTGGCCAGATGCTTATTCAATTGGCAAAAAAATAAAACGGTTTCAGCAAG 23414
```

```
Qy 595 ATTGTAACCTGTTGCTTTCTAAAGCATGAAAGCTTTTAAAGTCTTATGCTGCTGATGAT 654
Db 23413 ATCATCGTGTGCTTCTCGTAAACATGAAATAATCTGTTGAAAGAGTACGCTGCAGATGA 23354

Qy 655 GTCTTTGACTATCATGATCAGCGCTTATAGCAGATCAATTCGAAGTATCCAAACCTG 714
Db 23353 CTTTGTGACTACCAAGTCTGACGTTTATCGAAGATGAAAGAGTACAAACATTT 23294

Qy 715 CAACATGTTTATTCACCGCTGTGGGAAGAGATAGTATCCCCGAGCCCTATAAAGTCACA 774
Db 23293 CTTTACTTGTGTGACTGTGTCTCCACACAGAAACTATTCAACAGGTGTACAAATGTGCC 23234

Qy 775 GCAGATGATCTACCTCCACATTTATTAGAGTGGTTTCCAATGACCAATTCGAAGCATTCCT 834
Db 23233 GCTGATGACTTAGACGCTACGCTCGTTCAATTGACCGTTTAAACCGAAGATATCAAG 23174

Qy 835 GAAGAAATCAGAAAGATTAATGTTAAATTTGATATTTACTTTGTTGTTGCTGCTGCTGCT 894
Db 23173 GAGGAAGACAGGAGGCAAAACGTCAGTATTGAAGGAACCTCTCTATTTATGATAGGAGT 23114

Qy 895 CAAGAAATTTCTATTGGGTGCAACAGATTTCCTGTAGTCCAGATATCATGAAGCCACA 954
Db 23113 AACAGCTCCATTGCGACGTTTACTTTCGACAGAGACCTGGAATACAGGAAGCGGCC 23054

Qy 955 GTTAAATTCGTTAAGTTTATAATCCACACCTTTAAACAAGCGTGATATCCATCATATGAAT 1014
Db 23053 ATAAATTTATTAACTTCAATCAATCAAAATCAATGATGGTGAATCCACCATCCCA 22994

Qy 1015 ATTAAGTTTTTCAGAACCGCTTAGATGATGTCACAGCTTCACTGAAGTATATAAAGAA 1074
Db 22993 GTCAAGTTTTTACAGAACCGGTTAGATGATATCCACAGTTACTTGTATGATATTAAGC 22934

Qy 1075 GGTAAACCAAAATCTTAAGTATGTTGCCAGGTTATAA 1113
Db 22933 GGGAGGAATTCGCGCAAAAGTTGGTTCGCTTGTGAAA 22895
```

#### RESULT 13

##### SCCHRII/c

LOCUS 316613 bp DNA linear PLN 29-JUN-2001  
DEFINITION S.cerevisiae chromosome III complete DNA sequence.  
ACCESSION X59720.2 S43845 S49180 S50884 S93798  
VERSION X59720.2 GI:14588895  
KEYWORDS chromosome.  
SOURCE Saccharomyces cerevisiae (baker's yeast)  
ORGANISM Saccharomyces cerevisiae  
Bukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
Saccharomycetales; Saccharomycetaceae; Saccharomyces.

##### REFERENCE

- 1 Rad, M.R., Lutzenkirchen, K., Xu, G., Kleinhans, U. and Hollenberg, C.P.  
The complete sequence of a 11,953 bp fragment from ClG on  
chromosome III encompasses four new open reading frames  
Yeast 7 (5), 533-538 (1991)  
JOURNAL MEDLINE  
PUBMED 1897318  
REFERENCE 1
- 2 Biteau, N., Frenaux, C., Hebrard, S., Menara, A., Aigle, M. and  
Crouzet, M.  
The complete sequence of a 10.8kb fragment to the right of the  
chromosome III centromere of Saccharomyces cerevisiae  
Yeast 8 (1), 61-70 (1992)  
JOURNAL MEDLINE  
PUBMED 1580102  
REFERENCE 2
- 3 Wilson, C., Bergantino, E., Lanfranchi, G., Valie, G., Carignani, G. and  
Frontali, L.  
A putative serine/threonine protein kinase gene on chromosome III  
of Saccharomyces cerevisiae  
Yeast 8 (1), 71-77 (1992)  
JOURNAL MEDLINE  
PUBMED 1580103  
REFERENCE 3
- 4







```
QY 22 AAAGCCGTCATCATTTGAAGGTGACAAAGCTGTTGTTAAACAGATGTCCTCAGTTCCAGAA 81
DB 10 AAAGCTGTCGTCATTTGAAGCGGTAAAGCGGTGTCGAAGGCGGTTCCTCATTCCTGAA 69
QY 82 TTAAGAGGGGTACAGCCCTTGGTGAAGGTTGAGGCTGTTCTGCTGTAACCCAACTGATGG 141
DB 70 TTGGAAGAAGGATTCGTAATTAAGACACACTCGCTGTTGCTGTAACCCGACTGATGG 129
QY 142 AAGCATATTCCTATATAGATTTGTCAGAGAGTTCAATTTCTAGATGTGACATTCGTGGT 201
DB 130 GCACACATTAAGTACAGAGGTCGGGCCCTCAAGGATCTATTCTGGATGTGACGTCGGGC 189
QY 202 ACAGTTGTCAACTGACCA---AATGCTAGTACTGACTTTGAAGGTTGGAGATACCGGT 258
DB 190 CAATTTGTCAAATTTGGGCCAGCGCTGATCCTTAAGACATTTTCTATTGTTGATATATT 249
QY 259 TTGGTTTTTTCACGCTGTTCCAAACAGATCTCTAAATAATGTTGATTCGTTGATAT 318
DB 250 TATGGGTTTCAATTCACGGATCTTCCGTAAGGTTTCCCTTCCAAATGGTGTCTTTGCTGAAT 309
QY 319 GCAGGGTTTATCCACCTTTGTTTACAA-----GAGTAACCTTAACCTCACTCACTGCT 372
DB 310 TCTGCTATTCAACTGTTGCTGCTACAAATCAACCAATCACTCAAAATTTTGGGTGAA 369
QY 373 GATGAAATTTCTGAAGGCCCTGTGAAGAACTTCGAATCTGCTGATCATTCATTCGCCAGTTTCG 432
DB 370 GATGTTCTACTCGCGGCCCTGTGAGTCTTTGGAGGGGCGAGCCACTATCCAGTGTCA 429
QY 433 TTGACAACTGCTGTTGTTGTTGTTGTCATCCTTGGGCTCAAAATTTGGAATGGCAACCA 492
DB 430 CTGACCAACAGCTGGCTGTTGTTGACCTATAAATTTGGGCTTGAACCTGAAAGTGGAGCCA 489
QY 493 TCTACCCCGCAACATCTCATCTATTTGATTTGTTGGGTTGTTGCTGATCAGCAGTGGTCAA 552
DB 490 TCAACCCCAACAAAGAAACGGCCCATCTTATTATGGGGCGGTGCAACTGCAGTAGTGCAG 549
QY 553 CAACATATCAAGTTGCAACATATCAATGCTTATATAAGATTTGAATGTTGCTTCT 612
DB 550 TCGCTATCAATTAGCAATTAATGAAATGCTTCAACAGATCATTTGTTGGCTTCT 609
QY 613 AAAAGCATGAAAGCTTTTAAAGTCTTATGGTCTGATGATGCTTTTGAATCATGAT 672
DB 610 CGGAAACACGAAACCTGTTGAAAGATATGGTCTGATCAACTATTATTGATTACCATGAT 669
QY 673 GCAGGCTTATGAGCAGATCAATCGAAGTATCCAAACCTGCAACATGTTATTGACGCT 732
DB 670 ATTGACGTGTGAAACAATTTAAACACAGGTACAAATATCTCGTATTATTAGTCGACTGT 729
QY 733 GTGGGAAGCGAAGATAGTATCCCGAGGCTTATAAGTCAAGCAGATAGTCTACCTGCC 792
DB 730 GTCGGAAATCAAAATACGTTTCAACAGTGTACAAATGTGCGGCCGATTAACAGGATGCT 789
QY 793 ACATATTATGAAAGTGGTTCCAAATGACCATTTGAAAGCATTTCTGAAAGAAATCAGAAAGAT 852
DB 790 ACCGTTGCGAAATTAACATAATTTGACAGAAAGAAACGTCAAAAGAGAGATAGAGGCAA 849
QY 853 AATGTTAAATTTGATATTACTTTGTTATCGTGATCTCGTCAAGAAATTTCTATTGGGT 912
DB 850 AATGTCTATTGACAGAAAGAACTGTTTCAATAGGCGGCCATGAAGTACCAATTTGGT 909
QY 913 GCAACAGATTTCTGCTAGTTCAGAAATATCATGAAGCCACAGTTAAATTCGTTAAGTTT 972
DB 910 GGCATTAATTTCTGCTGATCCCAAGAGCCAGGAGAGCTGCCACGAAATTCGTCAGTTTC 969
QY 973 ATAAATCCACACCTTAAACAGGTTGATATCCATCATATGAATATTAAGTTTTCAGCAAC 1032
DB 970 ATCAATCCAAAGATTAGTATGGGCAAAATTCACCATATTCAGCAAGGGTCTATAGAAC 1029
QY 1033 GGCTTAGATGATGTCACGCTCTCAGTGAAGGTTATAAAGAGGTTAAACAAATATGTT 1092
DB 1030 GGGCTTTACGATGTTCTCGTATCTCTGGAAGACATTTAAATTCGTTAAGAACTCTGGTGA 1089
QY 1093 AAGTATGTTGCCAGGTTATA 1112
```

Db 1090 AAACGTTGCGGTATATAA 1109

## RESULT 15

BD174424

LOCUS

DEFINITION

BD174424 1122 bp DNA linear PAT 18-MAR-2003  
Novel enone reductase, process for producing the same, and method  
for selectively reducing carbon-carbon double bonds of alpha,  
beta-unsaturated ketone using the same.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches 643; Conservative 0; Mismatches 448; Indels 9; Gaps 2;

QY 22 AAAGCGTTCATCATTTGAAGGTGACAAAGCTGTTGTTAAACAGATGTCCTCAGTTCCAGAA 81

DB 10 AAAGCTGTCGTCATTTGAAGCGGTAAAGCGGTGTCGAAGGCGGTTCCTCATTCCTGAA 69

QY 82 TTAAGAGGGGTACAGCCCTTGGTGAAGGTTGAGGCTGTTCTGCTGTAACCCAACTGATGG 141

DB 70 TTGGAAGAAGGATTCGTAATTAAGACACACTCGCTGTTGCTGTAACCCGACTGATGG 129

QY 142 AAGCATATTCCTATATAGATTTGTCAGAGAGTTCAATTTCTAGATGTGACATTCGTGGT 201

DB 130 GCACACATTAAGTACAGAGGTCGGGCCCTCAAGGATCTATTCTGGATGTGACGTCGGGC 189

QY 202 ACAGTTGTCAACTGACCA---AATGCTAGTACTGACTTTGAAGGTTGGAGATACCGGT 258

DB 190 CAATTTGTCAAATTTGGGCCAGCGCTGATCCTTAAGACATTTTCTATTGTTGATATATT 249

QY 259 TTGGTTTTTTCACGCTGTTCCAAACAGATCTCTAAATAATGTTGATTCGTTGATAT 318

DB 250 TATGGGTTTCAATTCACGGATCTTCCGTAAGGTTTCCCTTCCAAATGGTGTCTTTGCTGAAT 309

QY 319 GCAGGGTTTATCCACCTTTGTTTACAA-----GAGTAACCTTAACCTCACTCACTGCT 372



```

Db 310 TCTGCTATTTCAACTGTGGTTCCTACAAATCAACCAATGAACCTCAAAATTTTGGGTGAA 369
Qy 373 GATGAAATTTCTGAAGCCCTGTGAAGAACTTCGAATCTCTGTCATCATTCGCCAGTTTCG 432
Db 370 GATGTTCTACTCGCGCCCTGTGAGTCTTTGAGGGGAGCCACTATCCCAAGTGTCA 429
Qy 433 TTGACAACTGCTGTGTAGTTTGTGTATCATCAGCTTGGGCTCAAAAATGGAATGCAACCCA 492
Db 430 CTGACCAACAGCTGGCTTGGTGTGACCTATTAACCTTGGCTTGAACCTGAAGTGGAGCCA 489
Qy 493 TCTACCCGGCAACATCATCTCATTTATGATTTGGGGTGTGTACAGCAGTGGGTCAA 552
Db 490 TCAACCCCAACAAAGAAACGGCCCATCTTATTTATGGGGCGGTGCAACTGCAAGTAGGTGAG 549
Qy 553 CAACATATCAAGTTGCGCAACATATCAATGCTTTATCTAAGATTGTAACCTGTGTCTTCT 612
Db 550 TCGCTATCTCAATTAGCAATAAATGAATGGCTTCCACCAAGATCATTTGTGTGGCTTCT 609
Qy 613 AAAAAGCATGAAAGCTTTTAAAGTCTTATGGTCTGATGATGTCTTTTGACTATCATGAT 672
Db 610 CGGAACACGAAAGCTTTCGAAGATATGGTCTGATCAACTATTTGATTACCATGAT 669
Qy 673 GCAGCGCTTATGAGCAGATCAATCGAAGTATCCAAACCTGCAACATGTTATTGACGCT 732
Db 670 ATTGACGTGGTAGAACAATTTAAACACAAGTACAACAATATCTCGTATTGATCGACTGT 729
Qy 733 GTGGGAAGCAGATAGTATCCCGAGGCTTATAAGTCAAGCAGATAGTCTACCTGCC 792
Db 730 GTCCGGAATCAAAATACGCTTTCAACAAGTGTACAAATGTGCGGCCGATAAACAGGATGCT 789
Qy 793 ACATTATTAGAAAGTGGTTCCAAATGACCAATGAAAGCAATCTCTGAAGAAATCAGAAAAAGAT 852
Db 790 ACCGTTCTGAAATTAATTTGACAGAGAAACGTCGAAAGGAGATAGAGGCA 849
Qy 853 AATGTTAAATGATATTAATTTGATGTCGATCTGTCGATCAAGAAATTTCTATTGGGT 912
Db 850 AATGCTACTATTGACAGAACAAAGCTGATTTCAATAGCGGCCATGAAGTACCAATTTGGT 909
Qy 913 GCACAGAGATTTCTGCTAGTCCAGATATCATGAAGCCACAGTTAAATTCGTTAAGTTT 972
Db 910 GGCATTACTTTCTGCTGACCCAGAGAGAGCTGCCACCGAATTCGTCAAGTTC 969
Qy 973 ATAATCCACACCTTTAAACAGCGTATATCATATGATATTAAGTTTTCAGCAAC 1032
Db 970 ATCAATCCAAAGATTAGTGATGGCAATTCACCATATTCGACGAAGGCTCTAAGAAC 1029
Qy 1033 GGCTTAGATGATGCTCCAGCTCTCACTGAAGGTATAAAGAGGTAAACCAAAATGTT 1092
Db 1030 GGGCTTTACGATGTTCTCGTATCTCGAGACATATTAATTCGTTAAGAACTCTGGTGAA 1089
Qy 1093 AAGTATGTTGCCAGTTATA 1112
Db 1090 AAACGTCGCGGATTATA 1109

```

## RESULT 16

CNS06FEH 1083 bp DNA linear STS 10-JAN-2001  
 LOCUS T7 end of clone AR0AA029E02 of library AR0AA from strain CBS 732 of  
 DEFINITION Zygosaccharomyces rouxii, sequence tagged site.

ACCESSION AL396303

VERSION AL396303.1 GI:12148292

KEYWORDS STS.

SOURCE Zygosaccharomyces rouxii

ORGANISM Zygosaccharomyces rouxii

REFERENCE Souciet, J.-L., Aigle, M., Artiguenave, F., Blandin, G.,

Boilotin-Fukuhara, M., Bon, E., Brottier, P., Casaregola, S.,

de-Montigny, J., Dujon, B., Durrens, P., Lepingle, A., Llorente, B.,

Malpertuy, A., Neuveglise, C., Ozier-Kalogeropoulos, O., Potier, S.,

Saurin, W., Tekala, P., Toffano-Nioche, C., Wesolowski-Louvel, M.,  
 Wincker, P. and Weissenbach, J.  
 Genomic exploration of the hemiascomycetous yeasts: 1. A set of  
 yeast species for molecular evolution studies  
 FEBS Lett. 487 (1), 3-12 (2000)  
 20584711  
 11152876  
 2 (bases 1 to 1083)  
 de Montigny, J., Straub, M., Potier, S., Tekala, P., Dujon, B.,  
 Wincker, P., Artiguenave, P. and Souciet, J.  
 Genomic exploration of the hemiascomycetous yeasts: 8.  
 Zygosaccharomyces rouxii  
 FEBS Lett. 487 (1), 52-55 (2000)  
 20584718  
 11152883  
 3 (bases 1 to 1083)  
 Genoscope.  
 Direct Submission  
 Submitted (06-SEP-2000) Genoscope - Centre National de Sequencage,  
 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :  
 seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)  
 This STS is part of a random genomic sequencing program of thirteen  
 yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces  
 exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,  
 Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces  
 lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia  
 angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,  
 Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to  
 5 kb were prepared and both extremities were sequenced. See  
 keywords for description of this sequence and for the sequence of  
 the other extremity of this insert.

FEATURES  
 Location/Qualifiers  
 1..1083  
 /organism="Zygosaccharomyces rouxii"  
 /mol\_type="genomic DNA"  
 /strain="CBS 732"  
 /db\_xref="taxon:4956"  
 /clone\_lib="AR0AA029E02"  
 /clone\_lib="AR0AA"  
 /notes="end : T7"  
 <10..>1008  
 /notes="similar to Saccharomyces cerevisiae ORF YLR460c [  
 similarity to C.carbonum toxD protein ]  
 similar to Saccharomyces cerevisiae ORF YNL134c [  
 similarity to C.carbonum toxD gene ]"  
 /evidence=not\_experimental  
 misc\_feature  
 <34..>1008  
 /notes="similar to Saccharomyces cerevisiae ORF YCR102c [  
 similarity to C.carbonum toxD gene ]"  
 /evidence=not\_experimental  
 BASE COUNT 330 a 206 c 239 g 306 t 2 others  
 ORIGIN

Query Match 22.7%; Score 252.4; DB 11; Length 1083;  
 Best Local Similarity 55.7%; Pred. No. 6.7e-46;  
 Matches 596; Conservative 2; Mismatches 443; Indels 29; Gaps 5;  
 Qy 1 ATGTGAGTTCACCACTCAAAAAGCGCTCATATTGAAGGTGACAAAGCTGTGTTAAA 60  
 Db 16 ATGTCAATTCCAACTAGTATGAAGCTGCTCATTTGAAGCGGTAAACGCTGCTGAAG 75  
 Qy 61 ACAGATGTCTCAGTTCAGAAATTAAGGAGGGGTACAGCCCTTGGTGAAGGTTGAGGCTGTT 120  
 Db 76 GAAGAAGTTCCTTTTACCACCTCTCAGGAGGGTCAAGTGTGTAGTCAAGAGCGTTGAGCT 135  
 Qy 121 GCTGGTAACCAACTGATTGGAAAGCATATTGCTTATAAGATTGGTCCAGAAAGTTCAATT 180  
 Db 136 GCTGTTAATCCACGGATTGGAAGCATATTGAGTATATAAACTTGTGCCACAAAGCTCAATT 195  
 Qy 181 CTAGGATGTGACATTTGCTGTGACAGTTGTCAAACCTTGGACCAAAATGCTAGTACTGACTTG 240  
 Db 196 TCAGGATGTGACGCTGCGGACAAAGTTGTAAGATTAGGGGCTAACCGTTAAATAATGTCGCT 255



```
Db 755 AATTGATCAAAATTAAGACCAAGTACACTGATTTTGTGTTTATAGACTGTGCTCCAC 696
Qy 741 CGAAGTAGTATCCCGAGGCTTATAAGTCACAGCAGATAGTCTACCTGCCACATTATT 800
Db 695 TCTAGACACTTTCAACCAAGTGTACCAATGTGCTCCACCTGATCGTAAATCAACTGTCTAT 636
Qy 801 AGAAGTGGTTCGAATGACCAATGGAAGCAATTCCTGGAAGAAATCAGAAAAGATA---ATGT 857
Db 635 CAATTATATGGGGTTTACTCTAGATGCTATTAAACCAAGATTAAAGACAACAGAAAGATGT 576
Qy 858 TAAATTTGATATTACTTTGTTGTTGATCTGTCATCTGTCAGAAATTTCTTATGGGTGCAAC 917
Db 575 CTCCTTCGAGGACACTGCTCTTACACTTGTGTTAGTGTTCATKTTAGTGAATA 516
Qy 918 AAGATTTCTGCTAGTCCAGATATCATGAAGCCACAGTTAAATTCGTTAAGTTTATAA 977
Db 515 CACAGTCAAGCTGANGCAAGANACAGACAGANNANCATBAANATGTGGAAATTAATAA 456
Qy 978 TCACACACTTTAACACCGGTGATATCCATCATATGAATATTAAAGTTTTTCAGCAACGGCTT 1037
Db 455 TGATAAATTCACCAATGGTTTCATTGACCAATATCCCAATTAAGTGTACAAAATGGGTT 396
Qy 1038 AGATGATGTCCTGAGCTCTCAGAGGTATATAAGAAAGGTAAACAAATAATGTTAAGTA 1097
Db 395 AGAATCTACCAATTCAAATCATATAATGATTTGGAATAATGGTAAACACTCTGCTGAAAAATT 336
Qy 1098 TGTGGCCAGCTT 1109
Db 335 TGTGGCCACTKT 324
```

```
RESULT 18
YSCDC91A/c
LOCUS YSCDC91A 5355 bp DNA linear PLN 08-APR-1994
DEFINITION Saccharomyces cerevisiae cdc91 gene, complete cds.
ACCESSION L31649
VERSION L31649.1 GI:469465
KEYWORDS
SOURCE Saccharomyces cerevisiae (baker's yeast)
ORGANISM Saccharomyces cerevisiae
REFERENCE 1 (bases 1 to 5355)
AUTHORS Bi,E. and Pringle,J.R.
TITLE Sequencing and characterization of CDC91
JOURNAL Unpublished (1994)
COMMENT Original source text: Saccharomyces cerevisiae DNA.
FEATURES
    source
        1..5355
            /organism="Saccharomyces cerevisiae"
            /mol_type="genomic DNA"
            /db_xref="taxon:4932"
            2540..3724
            /gene="cdc91"
            2540..3724
            /gene="cdc91"
            /codon_start=1
            /protein_id="AAA34487.1"
            /db_xref="GI:469466"
            /translation="MDSALKVALGCTAIRLAVNSLPFSLQQLDQSFSTPTVTSFR
            SLOEGLVLRNNIQVNVHGVHPPIIFPLSLFNDRLLIYLALIDGLIAYQLTEV
            TKAQNLKLVKPLGLVAVNPLTLLSIRSSIFITNPAISLSLVCILAEGNVLLSS
            VMLISGLSVPTLLIPLGLMKSWRQILNAILVLSLLILLPSYLSILQSWS
            FLQVGSIIPTFKVFNGLGWYFFLEMDTIFPPKAVFNIFIAVITPFTLRHK
            QPFYFLVLCIIVGTPKPSLDGAGFFPLPFTPLFGLYRPIISALLFLHAILV
            APIYHLWVILGSGNSNFFYAIISLVALASILVLDLWMLRIEYDNGIPNFKLKYT
            QI"
BASE COUNT 1621 a 975 c 933 g 1826 t
ORIGIN
Query Match 9.3%; Score 103.2; DB 8; Length 5355;
Best Local Similarity 54.1%; Pred. No. 1.2e-12;
```

```
Matches 210; Conservative 0; Mismatches 178; Indels 0; Gaps 0;
Qy 725 TTGACCTCTGGGAGCGAAGTAGTATCCCGAGGCTTATAAGTTCACAGCAGATAGTC 784
Db 5354 TCGACTGTGTGCGGAATCAAGTACCGTTCAACAGTGTACAAATGTGCGGCCGATAAAC 5295
Qy 785 TACCTGCCACATTATTAGAAGTGGTTCCCAATGACCATTGAAAGCATTCCTGAAGAAATCA 844
Db 5294 AGGATGCTACAATTTGTGAATTAATAAAATTTGCAGAGAGAAACGTCAAAAGAGAGAAC 5235
Qy 845 GAAAAGATTAATGTTAAATTTGATATTACTTTGTTGATCGTCATCTGCTCAAGAAATTC 904
Db 5234 GGAGACAAACGTTACTATTGACATAAATAGGCTATATTCAATAGTGGCCATGAAGTAC 5175
Qy 905 TATTGGGTGCAACAAGATTTCCTGTAGTCCAGAATATCATGAAGCCACAGTTAAATTCG 964
Db 5174 CATTTGGAACATTACTTTACCAGCGACTCAGAGCTAGGAAGCTGCAATAAATTTTA 5115
Qy 965 TTAAGTTTATAATCCACACCTTAAACACGGTGTATATCCATCATATATGAATATTAAAGTTT 1024
Db 5114 TCAAAATTCATCAATCCAAAGATTAAATGATGGACAAATTCGCCATATTCAGTAAGGGTCT 5055
Qy 1025 TCAGCAACCGCTTAGATGATGTCCTCCAGCTCTCACTGAAGGTATAAAGAGGTAAAAACA 1084
Db 5054 ATAAGAACCGGCTTTGTGATGTTCTCTCATATCTTAAAGACATCAATATGTTAAGAACT 4995
Qy 1085 AAAATGTTAAGTATGTTGCCAGGTTATA 1112
Db 4994 CTGGTGAATAAATCTGTTGCCGTATTAA 4967

RESULT 19
AX536956
LOCUS AX536956 1086 bp DNA linear PAT 22-NOV-2002
DEFINITION Sequence 557 from Patent WO202064766.
ACCESSION AX536956
VERSION AX536956.1 GI:25263391
KEYWORDS
SOURCE Candida albicans
ORGANISM Candida albicans
REFERENCE 1
AUTHORS Contreras,R.H., Eberhardt,I., Luyten,W.H. and Reekmans,R.J.
TITLE Bax-responsive genes for drug target identification in yeast and fungi
JOURNAL Patent: WO 02064766-A 557 22-AUG-2002;
COMMENT JANSSEN PHARMACEUTICA N.V. (BE)
FEATURES
    source
        1..1086
            /organism="Candida albicans"
            /mol_type="genomic DNA"
            /db_xref="taxon:5476"
            359 a 174 c 223 g 330 t
BASE COUNT 359 a 174 c 223 g 330 t
ORIGIN
Query Match 6.4%; Score 71; DB 6; Length 1086;
Best Local Similarity 49.0%; Pred. No. 2e-05;
Matches 324; Conservative 0; Mismatches 310; Indels 27; Gaps 4;
Qy 35 TTGAAGTGACAAAGCTGTTGTTTAAACAGATGTCTCAGTTCAGAAATTAAGAGGGTA 94
Db 29 TCGAACCTTATCAATTTAGCGGAAATTAAGATATTCCTCAACAAAAATAAAGAAATG 88
Qy 95 CAGCCTTGGTGAAGGTGTAGGCTGTGCTGGTGAACCACTGATTTGAAGACATATTGCTT 154
Db 89 AATATTAATCAAGCAGTAGTACTTTTGCATTAACCACTGATTTGAAGACATATTGTTT 148
Qy 155 ATAAGATTTGTCAGAGGTTCAATTTCTAGGATGTGACATTTGCTGGTACAGTTGTCAAAC 214
Db 149 ATCAATTTGGCAGCCCGAGGTGATGTTGTTGGTTCGATTTAGTGGATCAITTAAGAAG 208
Qy 215 TTGGACCAATGTAGTACTGACTTGAAGGTTGAGATACCGGTTTTCGGTTTTTTCACG 274
```

```
Db 209 TGGTCTCTCAAG---TAATCGGGTTGCAAAAGGTGACACTGTAAAGTCTTTTATAACTG 265
Qy 275 GTGCTTCCCAACAGATCTTAAATGGTCAATTTGCTGAATATGATGCGAGGTTTATCCAC 334
Db 266 GTA-----ATAGATCACTCGCACTGGAGCTTTTGCAGAAATATGTAGCTTGTGATCCTG 319
Qy 335 CTTTGTGTTTCAACAGAGTAACCTTAACCTCACTCAACTGCTGATGAAATTTCTGAAGGCCCTG 394
Db 320 CTACTTCGATAAAGTACATAAGAAATTTTGAACATTTGACTAAATTAACAAGTATCTGAAA 379
Qy 395 TGAAGAACTTTCGAATCTGCTGCATCAATGCGCAGTTTCGTTGACAACTGCTGGTGTAGTT 454
Db 380 TCCACTCATTTGAAGGGGACGAGTATTAATTTAGGTTTGGTTTACCGTTGGGCTTTTCAT 439
Qy 455 TGTGTCACTACTGGGCTCAAAATGGAATGGAATGGACCACTACCCCGCAACATACTCATC 514
Db 440 TTTCTCTAATCTACGAATTTGACAAACAAAAGCAACC-----TGGGGATA 484
Qy 515 CATTATTGATTGGGTTGGTCTACAGCAGTGGGTCAACAACTAATCCAAAGTTGCCAAAC 574
Db 485 GTATTTTGAATTTGGGAGGAGCAACTGCACTGAGTTCTAGCAATTCAGGTTGCCAAAC 544
Qy 575 ATATCAATGCTTATCAAGATTTGAATGTTTAACTGTTGCTTCTAAAAGCATGAAAAGCTTTAA 634
Db 545 TAGTGA--TAATCTCAAAGTAATCAACAGCATCACCAACAAACACACACCTCTTGA 601
Qy 635 AGCTTATGCTGATGATGCTTTGACTATCATGATGAGGCTTATGAGCAGATCA 694
Db 602 AACAAATTAGGGGACAGATTATTTTTCGATTATGAGACGCTGATGTTGCAATAAATTA 661
Qy 695 A 695
Db 662 A 662
```

```
RESULT 20
CCTOXD C.carbonum toxD gene.
LOCUS C.carbonum toxD gene.
DEFINITION X92391
ACCESSION X92391.1 GI:1027510
VERSION toxD gene.
KEYWORDS Cochliobolus carbonum
SOURCE Cochliobolus carbonum
ORGANISM Cochliobolus carbonum
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Dothideomycetes;
Pleosporales; Pleosporaceae; Cochliobolus.
REFERENCE 1 (bases 1 to 1506)
AUTHORS Cheng, Y.Q., Ahn, J.H. and Walton, J.D.
TITLE ToxD, a gene cloned from a HC-toxin producing strain of
Cochliobolus carbonum, has no role in HC-toxin biosynthesis
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1506)
AUTHORS Cheng, Y.
TITLE Direct Submission
JOURNAL Submitted (09-OCT-1995) Y. Cheng, Michigan State University,
MSU-DOE Plant Research Laboratory, Room 210 Plant Biology Building,
East Lansing, Michigan 48824-1312, USA
REFERENCE 3 (bases 1 to 1506)
AUTHORS Walton, J.D.
TITLE Direct Submission
JOURNAL Submitted (09-OCT-1995) J.D. Walton, Michigan State University,
MSU-DOE Plant Research Laboratory, Room 210 Plant Biology Building,
East Lansing, Michigan 48824-1312, USA
FEATURES
source
1. .1506
/mol_type="genomic DNA"
/db_xref="taxon:5017"
/chromosome="largest (3.5 Mb)"
289. .1182
/gene="toxID"
289. .1182
gene
CDS
```

```
/gene="toxID"
/function="unknown"
/notes="unique to isolates that make the cyclic peptide
HC-toxin"
/codon_start=1
/protein_id="CAA63129.1"
/db_xref="GI:1027511"
/db_xref="SWISS-PROT:P54006"
/translation="MTFQKAVTEAPHRARLYSDRLIPKLRDDYILVTVSVVALNPTD
WHHLKSPGCLVGCYAGIVVEVGRSVKPKFGDRVCGFAGHGNATVFDGDTAE
VITVKGDIQAWIPENLSFOEAATLVGKIVTGGVGLQSLKSWPPTPIHVAIPILY
GSTATGTLAIQAKLSGYRVTTCSPHFELMKSLGADLVFDYHEITTSADHRRCTQN
KKLKVPDTISIDVSARFCDRAMTEGGEYSALLDVSIAINTISSRWTAYTVLGEQFT
SQIVFPQPYPTMTSSGRNFGM"
polyA_site 1382
BASE COUNT 399 a 316 c 361 g 430 t
ORIGIN
Query Match 4.8%; Score 53.2; DB 8; Length 1506;
Best Local Similarity 54.1%; Pred. No. 0.19;
Matches 131; Conservative 0; Mismatches 108; Indels 3; Gaps 1;
Qy 74 TTCAGAAATTAAGGAGGTACAGCTTGGTGAAGCTTGAGGCTGTGCTGTAACCCAA 133
Db 356 TTCCCAAGTTGAGAGACGATTACATACTCGTCAGGACGGTAAGTGTGCTCTTAACCCCA 415
Qy 134 CTGATTGGAAGCATATTTGCTTATAAGATTGGTCCAGAAAGTTCAATCTAGGATGACA 193
Db 416 CTGATTGGAAGCATATTTTTCGCT---CTCTACCCCGGCTGTCTGGTGGCTGGCACT 472
Qy 194 TTGCTCGTACAGTTGTCAAACTTGGACCAATGCTAGTACTGACTTGAAGTTGGAGATA 253
Db 473 ACCTGGCATCGTTGAGGAAGTAGGACGAGGTGTTAAAACCGTTCAAAAAGGTGATC 532
Qy 254 CCGGTTTCGGTTTTGTTCAAGTGTCTCCCAACAGATCTTAAAAATGTCGATTTGCTG 313
Db 533 GAGTATCGGCTTTGCTCAGGGGGTAATGTCAGTTTTTTTCAGATGATGAACATTTGCCG 592
Qy 314 AA 315
Db 593 AA 594
RESULT 21
AC136676/c
LOCUS AC136676
DEFINITION Rattus norvegicus clone CH230-141C19, *** SEQUENCING IN PROGRESS
***, 2 unordered pieces.
ACCESSION AC136676
VERSION AC136676.3 GI:30521586
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 212955)
AUTHORS Muzny, D.Marie., Metzker, M.Lee., Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Ayodeji, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
Chacko, J., Chavez, D., Chen, R., Chen, Y., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Delgado, O., Denson, S., Davy-Carroll, L., De Anda, C., Dederich, D.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
```

Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Haves, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hoques, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowals, C., Krafc, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuhs, L., Loulseghe, H., Lozado, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhiney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J.J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokemele, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., Fuao, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajls, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wlezyk, R., Woodden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

Direct Submission  
Unpublished  
2 (bases 1 to 212955)  
Rat Genome Sequencing Consortium.  
Direct Submission  
Submitted (07-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 212955)  
Rat Genome Sequencing Consortium.  
Direct Submission  
Submitted (10-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On May 10, 2003 this sequence version replaced gi:25188698. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
----- Project Information  
Center project name: GHKS  
Center clone name: CH230-141C19  
----- Summary Statistics  
Assembly program: Atlas 3.0;  
Consensus quality: 200763 bases at least Q40  
Consensus quality: 203589 bases at least Q30  
Consensus quality: 205400 bases at least Q20

Estimated insert size: 210228; sum-of-contigs estimation  
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation  
-----  
\* NOTE: Estimated insert size may differ from sequence length (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently consists of 2 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.  
\* 1 158013: contig of 158013 bp in length  
\* 158014 158113: gap of unknown length  
\* 158114 212955: contig of 54842 bp in length.  
FEATURES  
source  
1. 212955  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-141C19"  
3356. 4203  
/note="clone boundary  
clone end:T7  
site:EcORI  
end sequence:RWAXI22TJC"  
3361. 4184  
/note="clone boundary  
clone end:T7  
site:EcORI  
end sequence:RWAXI22TJB"  
39881. 41507  
/note="wgs end\_extension  
clone end:T7"  
71654. 73136  
/note="wgs end\_extension  
clone end:T7"  
158114. 159255  
/note="wgs end\_extension  
clone end:T7"  
159306. 161347  
/note="wgs end\_extension  
clone end:T7"  
BASE COUNT 54541 a 47359 c 49948 g 54864 t 6243 others  
ORIGIN  
Query Match 4.4%; Score 48.6; DB 2; Length 212955;  
Best Local Similarity 47.5%; Pred. No. 1.5;  
Matches 144; Conservative 0; Mismatches 159; Indels 0; Gaps 0;  
QY 25 GCGTCATCATTCGAGGTGACAAAGCTGTGTGTTAAACAGATGCTCAGTTCAGATTA 84  
Db 156322 GCTGCTGTTATTGCGACTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTTT 156263  
QY 85 AAGAGGGGTACAGCCTTGGTGAAGGTTGAGGCTGTTGCTGTTAACTGATTCGAG 144  
Db 156262 GTTGATGGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 156203  
QY 145 CATATTGCTTATAAGATTGGTCCAGAGGTTCAATTCTAGGATGTGACATTCGTG 204  
Db 156202 CTTGCTGCTGCTTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 156143  
QY 205 GTTGTCAAACTGGACCAATGCTAGTACTGACTTGAAGTTGGAGATACCGGTTCC 264  
Db 156142 GATGTTGCTGCTATTCTGATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCT 156083  
QY 265 TTGTTTACGGTGGCTTTCCCAACAGATCCTAAAAATGGTGCAATTCGTAATG 324  
Db 156082 GTTGTTCTTGATGATGATGCTGTTGTTGCTGCTGCTGCTGCTGCTGCTGCT 156023  
QY 325 GTT 327  
|||



```

/gene="At2g44525"
/notes="synonym: F16B22.47; supported by full length cDNA:
Ceres:8402"
mRNA
  join(2606..2812,2969..3085,3370..3404,3546..3646,
3744..4019)
/gene="At2g44525"
  join(2640..2812,2969..3085,3370..3404,3546..3646,
3744..3830)
/gene="At2g44525"
  /codon_start=1
/product="expressed protein"
/protein_id="AAM14882.1"
/db_xref="GI:20197026"
/translaton="MAMRLKSMATLPKLIQSMRKEVPKHSNPNVLPSPRLRAFSLYDQIN
LIDNPBQDLRFQBNFTSFTVNGKVEGSLLCVGNLMSWSPKSEITFTDLSLIFO
TVRIPELLVGGORDHPVTPVVRQPVKSLGKMLKLETVDSRNAASTYINILNBSGRVA
AALLPVGVTS"
gene
  4242..6380
/gene="At2g44530"
/notes="synonym: F16B22.2"
  join(4242..4337,4472..4648,4876..5124,5395..5538,
5646..5813,5885..5954,6035..6116,6206..6380)
/gene="At2g44530"
  join(4242..4337,4472..4648,4876..5124,5395..5538,
5646..5813,5885..5954,6035..6116,6206..6380)
/gene="At2g44530"
  /codon_start=1
/product="putative phosphoribosyl pyrophosphate
synthetase"
/protein_id="AAC27455.1"
/db_xref="GI:3341673"
/translaton="MASIVQSPFTFPALNRRSLRPPSSVFPFLCKNAADPYKFDG
GNAGPHLLGDTVPAGSFSTRLEDSIVQNTLRIRFSGTANPILAQEISCVLGLDGL
KIKRPADEGIVQLQESVRCGDVFLVQTPPANENMELLMWIDACRAARAGIT
AVIPFYGADRTQRESIAKLANLITQSGADRVLCADLHSGMSGYFIDIPVDH
VYQGPVLDLYAKVAIVDDIMDTAGTISKGALLHQEGAREVYACTTHAVFPFPAIS
RLSGLFQEVITNTIPLSEKIFPPQLTVLSVANLLGETIWRVHDDCSVSKT"
  complement(6742..6763)
  /rpt_family="AT_rich"
  /note="synonym: F16B22.3"
  complement(6851..7168,7280..7561,7648..8301,
8399..8620)
/gene="At2g44540"
  complement(join(6851..7168,7280..7561,7648..8301,
8399..8620))
/gene="At2g44540"
  /codon_start=1
/product="putative cellulase"
/protein_id="AAC27456.1"
/db_xref="GI:3341674"
/translaton="MMSGKNGSSQCLWTSICIVLVMSMARGAVSTNYABALQKSLLY
FEAQRSGKLPNQRTVWRGSDSALRGSDAHVDLTGGYDAGDNMKFGPPLAFTTTMLA
WSNIEMSELRAHEQNSRLKAWTDYLIKAHPQPNVLYQGVGEGNSDHKCMWRPE
DMTTPRTSYRIDAQHPGSDLAGETAAMAAASIAFAPDKAYANILIGHAKDLFAFAK
AHGILYQNSIPNAGGFYASSGYEDLLWAAWLHRAATNDQIYNLYLTQASGTGPRSV
FAMDKFKVAQVLVAKLVEGVKQEGKMAEYKSMAEQICNCAQKGNVVKTPGGL
LYFLPWNLYQYTTAATFVLSAYSKYLEAAKASIDCPDGAQASDLDLQARSQVDYILG
SNPQKSMYVGTGTYNPKPHRAASIVSIRKDKTPVTCSEGGYDKWYNNPAPNPVLM
GALVGGPNNDVYGDERSNFQQAEPATVTVAPFVGLAAVF"
  complement(9891..11652)
/gene="At2g44550"
/notes="synonym: F16B22.4"
  complement(join(9891..10208,10319..10600,10686..11336,
11431..11652))
/gene="At2g44550"
  complement(join(9891..10208,10319..10600,10686..11336,
11431..11652))
/gene="At2g44550"
  /codon_start=1
/product="putative glucanase"

```

```

/protein_id="AAC27457.1"
/db_xref="GI:3341675"
/translaton="MSQLKNGSSQCLWTSICIVLVMSMARGAVSTNYABALQKSLLY
FEAQRSGKLPNQRTVWRGSDSALRGSDAHVDLTGGYDAGDNMKFGPPLAFTTTMLA
WSNIEMSELRAHEKGNALRALKWATDYLIKAHPQPNVLYQGVGEGNSDHKCMWRPE
DMTTPRTSYRIDAQHPGSDLAGETAAMAAASIAFAPDKAYANILIGHAKDLFAFAK
AHGILYQNSIPNAGGFYASSGYEDLLWAAWLHRAATNDQIYNLYLTQASGTGPRSV
FAMDKFKVAQVLVAKLVEGVKQEGKMAEYKSMAEQICNCAQKGNVVKTPGGL
LYFLPWNLYQYTTAATFVLSAYSKYLEAAKASIDCPDGAQASDLDLQARSQVDYILG
SNPQKSMYVGTGTYNPKPHRAASIVSIRKDKTPVTCSEGGYDKWYNNPAPNPVLM
GALVGGPNNDVYGDERTDYQHAEPAPATAAPFVGLAAVA"
  complement(14607..14630)
  /rpt_family="AT_rich"
  /note="synonym: F16B22.5"
  complement(join(14630..15252,15445..15729,15807..16460,
16557..16778))
/gene="At2g44570"
/notes="synonym: F16B22.6"
  complement(join(14935..15252,15445..15729,15807..16460,
16557..16778))
/gene="At2g44570"
  /codon_start=1
/product="putative glucanase"
/protein_id="AAC27459.1"
/db_xref="GI:3341677"
/translaton="MSCISSQCFITIKSICIVLLLSITCGAVSANSNYGEALTKSLLY
FEAQRSGKLPDQRTVWRGSDSALRGSDAHVDLTGGYDAGDNMKFGPPLAFTTTMLA
WSNIEMSELRAHQEHGNTLAARWATDYLIKAHPQPNVLYQGVGDNMDHCKMWRPE
DMTTPRTSYRIDPQHPGDLAGETAAMAAASIAFAPDSYAKTLISHAKELFEPAK
FWDGFKFLAQNAGGFYSSGYEDLLWAAWLHRAATNDQIYNLYLTQASNSGARSV
FAMDKFKVAQVLVAKLVEGVKQEGKMAEYKSMAEQICNCAQKGNVVKTPGGL
LYFLPWNLYQYTTAATFVLSAYSKYLEAAQTSIQCPNGDVLQASDLDLQARSQVDYIL
GNPKKNSYVGTGTYNPKPHRAASIVSIRKDKTPVTCSEGGYDKWYNNPAPNPVLM
VGAIVGGPPEYDAYGDERSDFOHQGPEDTVTVAPFVGLAAIA"
  complement(16815..16914)
  /rpt_family="(TA)n"
  complement(join(17216..17237))
  /rpt_family="AT_rich"
  complement(18506..23922)
  /gene="At2g44580"

```

Query Match 4.3%; Score 48.2; DB 8; Length 109375;  
Best Local Similarity 50.7%; Pred. No. 1.9;  
Matches 116; Conservative 0; Mismatches 113; Indels 0; Gaps 0;

Qy 800 TAGAAGTGGTTCGAATGACCATTTGAAGCATTCCTCGAAGAAATCAGAAAGATAATGTTA 859  
|||||  
Db 19728 TAGTATATGAAACAACACTAAATTCGACATTATTGCTGCTGATTTGTTAAAGTTA 19787  
|||||

**KEYWORDS**  
**SOURCE**  
**ORGANISM**  
**REFERENCE**  
**AUTHORS**  
**TITLE**  
**JOURNAL**  
**MEDLINE**  
**PUBMED**  
**REFERENCE**  
**AUTHORS**  
**TITLE**  
**JOURNAL**  
**FEATURES**  
**Source**  
**gene**  
**CDS**  
**gene**  
**CDS**  
**gene**  
**CDS**

Clostridium acetobutylicum  
 Clostridium acetobutylicum  
 Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
 Clostridium  
 1 (bases 1 to 13844)  
 Nolling,J., Breton,G., Omelchenko,M.V., Markarova,K.S., Zeng,Q.,  
 Gibson,R., Lee,H.M., Dubois,J., Qiu,D., Hitti,J., Wolf,Y.T.,  
 Tatusov,R.L., Sabathe,F., Doucette-Stamm,L., Soucaille,P.,  
 Daly,M.J., Bennett,G.N., Koonin,E.V. and Smith,D.R.  
 Genome sequence and comparative analysis of the solvent-producing  
 bacterium Clostridium acetobutylicum  
 J. Bacteriol. 183 (16), 4823-4838 (2001)  
 21359325  
 11456286  
 2 (bases 1 to 13844)  
 Childress,D., Zeng,Q. and Smith,D.R.  
 Direct Submission  
 Submitted (24-JUL-2001) GTC Sequencing Center Phosphatase,  
 Finishing, and Bioinformatics Teams, Genome Therapeutics Corp., 100  
 Beaver Street, Waltham, MA 02453-8443, USA  
 Location/Qualifiers  
 1. .13844  
 /organism="Clostridium acetobutylicum"  
 /mol\_type="genomic DNA"  
 /strain="ATCC 824"  
 /db\_xref="ATCC:824"  
 /db\_xref="taxon:1488"  
 135. .1466  
 /gene="CAC0460"  
 135. .1466  
 /gene="CAC0460"  
 /codon\_start=1  
 /transl\_table=11  
 /product="CBS-domain containing protein, YHDP B.subtilis  
 ortholog"  
 /protein\_id="AAK78440.1"  
 /db\_xref="GI:15023318"  
 /translation="MEADPDPMILQFVLIAVLTLVNAPFSAEMSIVSLNTRMKI  
 LSEGNRAQILKELIKPTKELSTVQGITLASFSSNPAATGLSDMFGYLLKINI  
 PYSSQALIVTVISYFTLVFGLFPFKRIALQKSEKIAMLAARPILYVSKITVPFVK  
 LLSASTNVLIRITGSMNMNKLKSVIEHSKIVEAGQYGVSESEKEMIDSIFFD  
 RIAEYVPTRVYILINDTPLNKYVDELEIKYGRVPVFGDSNIIIGILYKDFIV  
 EARKYGVENPNTKIIHSPYFVPERKNINELFKELQAAKHMVLIIDEVGGFSGIVTI  
 EDLLIEWGNISDEVEDDNYIKKIDESTFIKGIPIRELNEQLDLDEETDDYDT  
 LGGLIKLILGHIPENEQSQSIKYKGIITFEIKVEKRIEKVEVHLQKDLLE"  
 1493. .1891  
 /gene="CAC0461"  
 1493. .1891  
 /gene="CAC0461"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Mercuric resistance operon regulatory protein,  
 MerR family"  
 /protein\_id="AAK78441.1"  
 /db\_xref="GI:15023319"  
 /translation="WRYYKIGLSKIVNLSTETIRYEEKIRLIKPDPENRYKKYSE  
 YIKRLKIIKANNYGFTLHESNPLSEVSTSSIEDILSSTIMNKVQOIEKECERLN  
 MYKNQLLKFFEEINDLQCPVLNRIINEKNT"  
 1955. .3151  
 /gene="CAC0462"  
 1955. .3151  
 /gene="CAC0462"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Protein of short-chain alcohol dehydrogenase  
 family"  
 /protein\_id="AAK78442.1"  
 /db\_xref="GI:15023320"  
 /translation="MIVKAKFVKGFIRDFVPGVCRREVLNQIDYCKKAIIGRFGPKVKL  
 IVGASGFLGATRIYVAGGPEAHITGVSYETGADRIRIGTAGWNNIIPFKEFAKKG  
 LVKNFIEDASKNKKYIKKIDFGKIDLFVYSLAARPKDKYKGNLYVEDCFSDKA  
 LGDREGPTIDVERDITIKVYSSASIIETETRKVMGCGDQEWCEELIYEDCFSDKA



TTIAYVIGSPRTYKIVREGTIGIAKQLEDKAKLNEKLNVRVIGGRAFVSVNKAALT  
KASAYITPFPYLAAILYWKKEKNIHENCIMERFSEKIIYSNEKIQFDDKGRLRMD  
DLERKQVQDEVDRIWSNITPENFELSVDYKGYKKEFNWNLNGFLDDGVDSKLDIEL  
LRKLEP"

gene complement (3180. .4226)

CDS /gene="CAC0463"

complement (3180. .4226)

/gene="CAC0463"

/codon\_start=1

/transl\_table=11

/product="Serine protease Do (heat-shock protein)"

/protein\_id="AAK78443.1"

/db\_xref="GI:15023321"

/translation="MKTHKLLSVLASVLLTSFTGCTFASNKKKITPDTANAQAST  
SVESPIKASKVLPVVGVTVTYIERDNTKKGVGSMIIDSQGYLLTNHVGAMTS  
KDLVSLVDSIGAKPLWANESLDLSIIKIDKNLOAVTLGDSKVDIGETAIGN  
PLGNFORTVSGIVASVNRVVEAGEVFMEDLLQTDASINPGSGPLDANGNVIG  
VNSAKISAEIGFAVINIVKPVLSKLTGQFKTPIGIGLDSKMGVNLNFKF  
GIYVINISPGSAAAGINKGDIILSVNGRKNINTMBLRESIYTIGANNTVSLKLT  
SGEKTVMVKIQSST"

gene 4351. .4692

CDS /gene="CAC0464"

4351. .4692

/gene="CAC0464"

/codon\_start=1

/transl\_table=11

/product="Protein, related to divalent cations, tolerance

CutA family"

/protein\_id="AAK78444.1"

/db\_xref="GI:15023322"

/translation="MQINQFKIPIPKYVEKLGDKLSEANVGKIGNYDHCWATNIV  
EGYWRPLEGSEPVGEIGKVCHECKREYVENALKIRKVHVPYEEPVINI  
PILNQLPENN"

gene complement (4841. .5632)

/gene="CAC0465"

complement (4841. .5632)

/gene="CAC0465"

/codon\_start=1

/transl\_table=11

/product="Predicted transcriptional regulator,

dica/hlpB/ansR family"

/protein\_id="AAK78445.1"

/db\_xref="GI:15023323"

/translation="WQDSNLEIKILKNKIYDTSISIFHNLKLLIQLEGTKKLAKSI  
GVSDLLSKYSGDAPFAETILYICKLYNITDEFNLPLTSSYPEDMLYDDTGPDD  
IFKODYFCYRVTVSGKSGKHESNLRTKNCIFRIYIKEGVKEFKGTYSKIDNLV  
PFNLYSREYGYAITMINKPSLNSKSIYGGIAMLSPDANSKPCCKIIISSKINRE  
LSPDILNNFLCFKSRGSGFNKISRYEDEVKDYFITNLNEETKL"

gene 5832. .6254

CDS /gene="CAC0466"

5832. .6254

/gene="CAC0466"

/codon\_start=1

/transl\_table=11

/product="Hypothetical protein"

/protein\_id="AAK78446.1"

/db\_xref="GI:15023324"

/translation="MNLKRVYDIKSLCEYEFQKLYFQDKIYGVYIYKNSNIEGRK  
YIEFWIITDFTGEINLLKIHDKLFNSKRGYVSLDDVGKVDLQMKPYYIYVEN  
GKLKKGYMNIIDYFTWLVKNKAVGKISPSIDSLKGEF"

gene 6297. .7106

CDS /gene="CAC0467"

6297. .7106

/gene="CAC0467"

/codon\_start=1

/transl\_table=11

/product="Uncharacterized membrane protein, homolog of

YDAH\_B\_subtilis"

/protein\_id="AAK78447.1"

/db\_xref="GI:15023325"

/translation="MNQIQLLILFNFIITPFIPTIGTLAYSTRLVGVKTKIAISFTVFN  
LTLVSRVATVQEBLLVYAEKNLYSSNLIVFNSIILVSGAASILGAVLIPTQRMF  
SKGMVFFSLEKSIPLRVVHSMTRRGIVSMKQCVVLPSRKSVEKENYKLPRIIVYNF

gene

CDS

ISVAALTVGALAPIYAFKIVPNLRATCVTLSSVVGIAQIQLMTIFIDPQMSIMTEBVM  
ERKCSSEDFKNCVLANVGVSKVAGTFAAIFILIPASFAIVFAKVGVDFISRI"  
7119. .7823

/gene="CAC0468"

7119. .7823

/gene="CAC0468"

/codon\_start=1

/transl\_table=11

/product="HAD superfamily hydrolase"

/protein\_id="AAK78448.1"

/db\_xref="GI:15023326"

/translation="MIKALIMDSKVLNKPATGEWIFPKPFYVVDWERNTIPLIKR  
KSAFQKANKYIEQKLIITTEBEEYKCFLEYIKLSKELYOLKLNNDIEITEDLVYN  
KPYDYKDAIYBIPKLSKYKLAIVSDAMPSELEGVFKAGLRKYFSSPVISSVKGVT  
KNEMMYKTALSELRLVPEBAVFIDDNIRNCEGAVKLGKISFVLSRKSRIYAYNKLFN  
RRTKSRNLKQLEEV"

gene 8009. .8584

/gene="CAC0469"

notes="spmA"

8009. .8584

/gene="CAC0469"

/codon\_start=1

/transl\_table=11

/product="Spore maturation protein A (gene spmA)"

/protein\_id="AAK78449.1"

/db\_xref="GI:15023327"

/translation="MINYIWFIFIFVTGIVLNGSTLSSSIINSTESSVKMLMISL  
LGIMCLMGVMRIAESKGLTDKLSIVLRPLKIVPKDISKNNVMGPMIMNITSNMLG  
LGNATPFGIKTVMENQKINKKSTATNDMAKPLVINAAQIQLLPTTIISIRACSSS  
NPAITVITPITVSVTVAIMGLIYCKILQRHP"

gene 8596. .9117

/gene="CAC0470"

notes="spMB"

8596. .9117

/gene="CAC0470"

/codon\_start=1

/transl\_table=11

/product="Spore maturation protein B (gene spMB)"

/protein\_id="AAK78450.1"

/db\_xref="GI:15023328"

/translation="MNYISSALIPILFISIVLYGMIKKVKYECFVEGAKDGIRICVN  
IPFYIAMLSSVAIFRESGAMOYLINILKPVFKIIGFPSELVPLIIVKPLSGSGALGV  
FABILKKYGADTVIGIINSGSTETIFTYITVYVYGAIKIKIRHTLWAAINADMTS  
IIVAILVVRPMGI"

gene 9426. .9998

/gene="CAC0471"

9426. .9998

/gene="CAC0471"

/codon\_start=1

Query Match 4.3%; Score 47.4; DB 1; Length 13844;

Best Local Similarity 52.8%; Pred. No. 3.3;

Matches 102; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

Qy 809 TTCCAATGACCATTTGAAGCATTCCTCGAAGAAATCGAAGAAATGATAAATGTTAAATGATA 868

Db 2163 TTGGAGTATCTTATGAACAGGAGCTACAGATAGAGATAGGAACAGCGGATGATATA 2222

Qy 869 TTACTTTTGTATCGTGCATCTGGTCAAGAAATTCCTATTGGGTGCACAAAGATTTTCCTG 928

Db 2223 ATAACATATTTTTTAAAGATTTTGTCTAAAAAAGAGATTTAGTTGCAAAAAAATTCATTG 2282

Qy 929 CTAGTCCAGATATCATGAGCCACAGTAAATTCGTTAAGTTTATAAATCCACACTTA 988

Db 2283 AGGATGCTTTTCTTAATGAACCAAGATAGTATTATTAAAGATATATAAGGATGAATTTG 2342

Qy 989 ACAACGCTGATAT 1001

Db 2343 GTAAATAGATTT 2355

RESULT 26

HSJ106901

LOCUS	HSJ106901	92357 bp	DNA	linear	PRI 04-APR-2000
DEFINITION	Human DNA sequence from clone RP5-106901 on chromosome 20. Contains part of the gene for a novel protein similar to a low density lipoprotein-related protein LRP16 and STSs, complete sequence.				
ACCESSION	AL049633				
VERSION	AL049633.3	GI:4775651			
KEYWORDS	HTG; LRP16.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1 (bases 1 to 92357)				
TITLE	Matthews, L.				
JOURNAL	Direct Submission				
COMMENT	Submitted (15-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone request: clonesrequest@sanger.ac.uk On May 11, 1999 this sequence accession replaced gi:4678553. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at <a href="http://www.sanger.ac.uk/Projects/C_elegans/wormpep">http://www.sanger.ac.uk/Projects/C_elegans/wormpep</a> This sequence is the entire insert of clone RP5-106901 The true left end of clone RP4-738116 is at 92256 in this sequence. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. This sequence was generated from part of bacterial clone contigs of human chromosome 20, constructed by the Sanger Centre Chromosome 20 Mapping Group. Further information can be found at <a href="http://www.sanger.ac.uk/HGP/Chr20">http://www.sanger.ac.uk/HGP/Chr20</a> RP5-106901 is from the library RPCI-5 constructed by the group of Pieter de Jong. For further details see <a href="http://www.chori.org/bacpac/home.htm">http://www.chori.org/bacpac/home.htm</a> VECTOR: pCYPAC2.				
FEATURES	Location/Qualifiers				
source	1..92357				
	/organism="Homo sapiens"				
	/mol_type="genomic DNA"				
	/db_xref="R2PD:RPCIP7040011069"				
	/db_xref="taxon:9606"				
	/chromosome="20"				
	/clone="RP5-106901"				
	/clone_lib="RPCI-5"				
repeat_region	445..535				
	/note="MIR repeat: matches 110..200 of consensus"				
repeat_region	708..763				
	/note="L2 repeat: matches 2653..2708 of consensus"				
misc_feature	1438..1602				
	/note="match: STS: Em:116330"				
repeat_region	1603..2171				
	/note="MLT2CA repeat: matches 6..500 of consensus"				
repeat_region	3983..4140				
	/note="L2 repeat: matches 2480..2735 of consensus"				
repeat_region	5015..5509				
	/note="LTR37B repeat: matches 1..468 of consensus"				
repeat_region	5707..5787				
	/note="MIR repeat: matches 66..154 of consensus"				
repeat_region	6118..6221				
	/note="LMC5 repeat: matches 7312..7409 of consensus"				
repeat_region	6401..6736				
	/note="LMC5 repeat: matches 7560..7913 of consensus"				
repeat_region	6803..6850				

```
/note="L1M43 repeat: matches 4167. .6304 of consensus"
29755. .30565
/note="L1M44 repeat: matches 5473. .6300 of consensus"
31368. .31840
/note="L1M37B repeat: matches 1. .468 of consensus"
31864. .33026
/note="T1ggr3b repeat: matches 7. .1228 of consensus"
33105. .33174
/note="L2 repeat: matches 2671. .2737 of consensus"
35600. .35623
/note="12 copies 2 mer aa 100% conserved"
35695. .35860
/note="M1R repeat: matches 23. .186 of consensus"
35966. .36119
/note="M1R repeat: matches 36. .198 of consensus"
37193. .37331
/note="MER58A repeat: matches 81. .224 of consensus"
37805. .37859
/note="M1R repeat: matches 49. .115 of consensus"
37957. .38255
/note="M1R1I repeat: matches 72. .410 of consensus"
39627. .39920
/note="L2 repeat: matches 2362. .2707 of consensus"
41385. .41697
/note="L1M16C repeat: matches 86. .387 of consensus"
41854. .42272
/note="L2 repeat: matches 1377. .1805 of consensus"
42557. .42851
/note="ALUSX repeat: matches 1. .295 of consensus"
43341. .43987
/note="L1M16B repeat: matches 5488. .6171 of consensus"
44255. .44294
/note="20 copies 2 mer ca 90% conserved"
44681. .45000
/note="MER47A repeat: matches 29. .364 of consensus"
46252. .46605
/note="MER63 repeat: matches 641. .988 of consensus"
46606. .46920
/note="Aluy repeat: matches 1. .304 of consensus"
46921. .46982
/note="MER63 repeat: matches 988. .1050 of consensus"
47017. .47040
/note="12 copies 2 mer tg 95% conserved"
47321. .47738
/note="L1R16A repeat: matches 16. .448 of consensus"
47944. .48244
/note="L1M4 repeat: matches 3552. .3871 of consensus"
48261. .48573
/note="L1M4C4 repeat: matches 7218. .7522 of consensus"
49230. .49257
/note="L2 repeat: matches 2626. .2653 of consensus"
49339. .49430
/note="M1R repeat: matches 137. .227 of consensus"
49468. .49525

Query Match 4.3%; Score 47.4; DB 9; Length 92357;
Best Local Similarity 52.8%; Pred. No. 2.9;
Matches 102; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 736 GGAAGCGAAGTAGTATCCCGAGGCTATAAAGTCACAGCAGATAGTCTACCTGCCACA 795
DB 6358 GAAAGCAGACAGAAACCTACAGATTAAAGGACAGATATCTTGATTCACAAACA 6417
QY 796 TTATTAGAGTGGTTCACCAATGACCATTTGAAGCATTCCTGAGAAATCAGAAAGATAAT 855
DB 6418 CTATGTAAAAAGTTACAAATTTAAAAATGAATAATACTTATTATATCAAGGAAAGAAA 6477
QY 856 GTTAAAAATTGATATTACTTTGTCATCGTCATCGTCAAGAAATTCCTATTGGGTGCA 915
DB 6478 TTTAATATTAAATTTTTCAGTGTGTAGTATTATTTTAACTATCTTATATTTTAGA 6537
QY 916 ACAAGATTTTCCTG 928
```

```
Db 6538 AATATATATACGG 6550

RESULT 27
AC008057 119082 bp DNA linear HTG 11-OCT-2000
LOCUS Homo sapiens chromosome 20 clone P1-10263, WORKING DRAFT SEQUENCE,
DEFINITION 3 unordered pieces.
AC008057
AC008057.4 GI:10765024
VERSION HTG; HTGS PHASE1; HTGS_DRAFT.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 119082)
AUTHORS Mammalia; Euthera; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Eukaryota; Metazoa; Primates; Catarrhini; Hominidae; Homo.
JOURNAL 1 (bases 1 to 119082)
AUTHORS McCombie, W.R.
TITLE Human Genomic Sequence, Chromosome 20
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 119082)
AUTHORS McCombie, W.R.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-1999) Lita Annenberg Hazen Genome Sequencing
COMMENT Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
Harbor, NY 11724, USA
On Oct 11, 2000 this sequence version replaced gi:9954568.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 68531: contig of 68531 bp in length
* 68532 102563: gap of unknown length
* 68532 102563: contig of 33932 bp in length
* 102564 102663: gap of unknown length
* 102664 119082: contig of 16419 bp in length.
FEATURES
Location/Qualifiers
source 1..119082
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="20"
/clone="P1-10263"
BASE COUNT 37455 a 22465 c 21882 g 36187 t 1093 others
ORIGIN
Query Match 4.3%; Score 47.4; DB 2; Length 119082;
Best Local Similarity 52.8%; Pred. No. 2.9;
Matches 102; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 736 GGAAGCGAAGTAGTATCCCGAGGCTATAAAGTCACAGCAGATAGTCTACCTGCCACA 795
DB 12709 GAAAGCAGACAGAAACCTACAGATTAAAGGACAGATATCTTGATTCACAAACA 12768
QY 796 TTATTAGAGTGGTTCACCAATGACCATTTGAAGCATTCCTGAGAAATCAGAAAGATAAT 855
DB 12769 CTATGTAAAAAGTTACAAATTTAAAAATGAATAATACTTATTATATCAAGGAAAGAAA 12828
QY 856 GTTAAAAATTGATATTACTTTGTCATCGTCATCGTCAAGAAATTCCTATTGGGTGCA 915
DB 12829 TTTAATATTAAATTTTTCAGTGTGTAGTATTATTTTAACTATCTTATATTTTAGA 12888
QY 916 ACAAGATTTTCCTG 928
DB 12889 AATATATATACGG 12901

RESULT 28
AC007630/c 169494 bp DNA linear PRI 22-NOV-2000
LOCUS AC007630
```





```
Db 162536 --AAATGTTGATAATGTAATGTTGATAATGTAATGTTGATAATGTTGATAATGTTGATAAT 162479
Qy 814 ATGACCATGGAAGCATTCCTGAAGAATCAGAAAGATAGTAATGTAATGTAATGTAATGTAAT 873
Db 162478 GTAGATGTTGATAATGTAATGTTGATAATGTAATGTAATGTTGATAATGTAATGTTGATAAT 162419
Qy 874 TTGTTGATCGTCATCTGGTCAAGAAATTCATTGGGTGCAACAAGATTTCCTGCTAGT 933
Db 162418 GTAATGTTGATAATGTAATGTTGATAATGTAATGTTGATAATGTTGATAATGTTGATAAT 162359
Qy 934 CCAGAATATCATGAAGCCACAGTAAATTCGTTAAATGTTTAAATCCACACCTTTAAACAAC 993
Db 162358 GTAAATGTTGATAATGTAATGTTGATAATGTAATGTAATGTTTAAATGAAGGTGATAAT 162299
Qy 994 GGTGATATCCATCATATGAATATTAAGTTTTCAGCAACGCTTAGATGATGTCCTCCAGCT 1053
Db 162298 ATTAGTGTAAATATGTTGATGCTCTTTATTAATATTTGGTGTTCATGATGATGACGATGAT 162239
Qy 1054 CTCACGAGGTATATAAAGAGGTAAACAAACAAAATGTTAAGTATGTTG 1102
Db 162238 GATCAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 162190
```

```
RESULT 30
AC096684 84118 bp DNA linear HTG 20-NOV-2002
LOCUS Takifugu rubripes clone 241N7, WORKING DRAFT SEQUENCE, 2 ordered
DEFINITION pieces.
AC096684
VERSION AC096684.2 GI:25139901
KEYWORDS HTG; HTGS PHASE2; HTGS DRAFT.
SOURCE Takifugu rubripes (Fugu rubripes)
ORGANISM Takifugu rubripes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodontidae; Takifugu.
1 (bases 1 to 84118)
Akhter,N., Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M.,
Benjamin,B., Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S.,
Carliaga,K., Coleman,B., Engle,J., Granite,S., Guan,X., Gupta,J.,
Haghighi,P., Han,J., Hansen,N., Ho,S.-L., Idol,J.R., Karlins,E.,
Laric,P., Lee-Lin,S.-Q., Legaspi,R., Maduro,Q.L., Maduro,V.B.,
Margulies,E.H., Mastello,C., Maskeri,B., McDowell,J.,
Pequirigan,C., Pearson,R., Portnoy,M.E., Prasad,A.,
Reddix-Dugue,N., Schandier,K., Schueler,M.G., Sison,C.,
Scantripop,S., Thomas,J.W., Thomas,P.J., Touchman,J.W., Vogt,J.L.,
Wetherby,K.D., Wiggins,L., Young,A. and Green,E.D.
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 84118)
Green,E.D.
Direct Submission
Submitted (22-SEP-2001) NIH Intramural Sequencing Center, 8717
GroveMont Circle, Gaithersburg, MD 20877, USA
3 (bases 1 to 84118)
Green,E.D.
Direct Submission
Submitted (20-NOV-2002) NIH Intramural Sequencing Center, 8717
GroveMont Circle, Gaithersburg, MD 20877, USA
On Nov 20, 2002 this sequence version replaced gi:15721942.
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc.zoo@nih.gov
----- Project Information
Center project name: arc
Center clone name: 241N07
```

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated

order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8X average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

----- Summary Statistics

Sequencing vector: plasmid; n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 83852 bases at least Q40  
Consensus quality: 83954 bases at least Q30  
Consensus quality: 83979 bases at least Q20  
Insert size: 89000; agarose-fp  
Insert size: 101000; pulse-field-gel  
Insert size: 84018; sum-of-contigs  
Quality coverage: 13.13x in Q20 bases; agarose-fp  
Quality coverage: 11.57x in Q20 bases; pulse-field-gel  
Quality coverage: 13.91x in Q20 bases; sum-of-contigs

-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 2 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submitter.

\* This sequence will be replaced

\* by the finished sequence as soon as it is available and

\* the accession number will be preserved.

\* 1 11009: contig of 11009 bp in length

\* 11010 11109: gap of unknown length

\* 11110 84118: contig of 73009 bp in length.

----- Location/Qualifiers

source

1. .84118

/organism="Takifugu rubripes"

/mol\_type="genomic DNA"

/db\_xref="taxon:31033"

/clone="241N7"

/clone\_lib="Incyte Genomics"

1. .11009

/note="assembly\_fragment"

clone\_end:17

vector\_side:left"

1. .7443

misc\_feature

/note="clone overlaps with GenBank Accession Number

AC096997 clone 217A23 (center project name arb)"

11110. .84118

/note="assembly\_fragment"

clone\_end:SP6

vector\_side:right"

BASE COUNT 20759 a 20909 c 21005 g 21345 t 100 others

ORIGIN

Query Match 4.2%; Score 47; DB 2; Length 84118;

Best Local Similarity 44.6%; Pred. No. 3.6;

Matches 185; Conservative 0; Mismatches 230; Indels 0; Gaps 0;

Qy 36 TGAAGTGCACAAAGCTGTTGTTAAACACATGCTCAGTTCAGATTCCAGATTAAAGAGGGTAC 95

Db 4887 TGTGTTGTTGCTGCTGCTGTTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4946

Qy 96 AGCTTTGGTGAAGGTTGAGGCTGTTGCTGTTAAACCACTGATTGAAGCATATTGCTTTA 155

Db 4947 TGTGTTGTTGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 5006

Qy 156 TAAGATTGCTCAGAAAGGTTCAATTCTAGATGTGACATGCTGTGACAGTTGTCTAACT 215

Db 5007 TTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5066



**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 10, 2004, 05:08:13 ; Search time 376 Seconds

(without alignments)

7990.624 Million cell updates/sec

Title: US-10-081-644-1

Perfect score: 1113

Sequence: 1 atgcagttcccaaccactca.....agtatgttccaggtataa 1113

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 90 summaries

Database : N Geneseq\_19Jun03.\*

```

1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*
25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	357.8	32.1	1631	24	ABQ76453
2	98.4	8.8	393	24	ABQ76679
3	71	6.4	1086	24	ABQ76566
4	64.4	5.8	1330	24	ABT11258
5	45.4	4.1	23439	18	AAV74349
6	40.8	3.7	3501	21	AAA70202
7	40.2	3.6	439	22	ABA49878
8	40.2	3.6	439	22	ABA67796
-----					
1	357.8	32.1	1631	24	ABQ76453
2	98.4	8.8	393	24	ABQ76679
3	71	6.4	1086	24	ABQ76566
4	64.4	5.8	1330	24	ABT11258
5	45.4	4.1	23439	18	AAV74349
6	40.8	3.7	3501	21	AAA70202
7	40.2	3.6	439	22	ABA49878
8	40.2	3.6	439	22	ABA67796

Probe #13320 for 9  
Human brain expres  
Human bone marrow  
Probe #12651 for 9  
Probe #12651 used  
Probe #16702 used  
Probe #8374 used  
Human liver single  
Human genome-deriv  
Human huntington (c  
AmBPV first RNA po  
Human immune/haema  
AmBPV genome fragm  
Strawberry fruit a  
Streptococcus poly  
Strawberry fruit a  
Streptococcus poly  
Human GDP-mannose  
Genomic sequence o  
Gastric cancer ass  
Polyglutamine trac  
Bacteriophage 96 c  
AmBPV genome fragm  
S. pneumoniae type  
L-lactic acid dehy  
Streptococcus pneu  
Buchnera sp. genom  
Streptococcus pneu  
Drosophila melanog  
Human canion gene  
Human cDNA clone (c  
DNA encoding novel  
Maize PCNA P120 ge  
Listeria innocua c  
Listeria innocua D  
Streptococcus poly  
Drosophila melanog  
Plasmodium falci  
Tumour suppressor  
Human flamingo CDN  
Human flamingo CDN  
Human ORFX ORF1956  
Human REPTX 9 CDNA  
Human FLAMINGO 1 h  
Human GPCR CELSR2  
Human mdt cDNA SE  
Human cDNA differe  
Gene #3817 used to  
Borrelia burgdorfe  
Streptococcus poly  
Human breast cell  
Human foetal liver  
Probe #1403 for ge  
Human brain expres  
Human bone marrow  
Probe #3426 for ge  
Probe #3537 used t  
Probe #3366 used t  
Human liver single  
Human genome-deriv  
Arabidopsis thalia  
Drosophila melanog  
Drosophila melanog  
Enterococcus faeca  
Drosophila melanog  
Drosophila melanog  
Enterococcus faeca  
Enteroecoccus faeca  
Methanococcus jann  
Sequence encoding  
Bacteriophage 3A c  
Arabidopsis thalia  
Nucleic acid seque  
Mycoplasma genital

C	82	36	3.2	22	21	ABN81097	Shrimp polynucleot
	83	36	3.2	22	23	AAH53374	Haemophilus influenzae
	84	36	3.2	978	22	AAH50794	CPE 98 coding seq
C	85	36	3.2	9095	24	ABK28448	DNA transcription
	86	36	3.2	40324	24	ABO67150	Human angiogenesis
C	87	36	3.2	1082138	21	AAF223305	Arabidopsis thaliana
	88	36	3.2	1830121	17	AA424063	Haemophilus influenzae
	89	35.8	3.2	840	20	AA221567	Staphylococcus aur
	90	35.8	3.2	840	20	AAK29297	S. aureus spo02 p

## ALIGNMENTS

RESULT 1

RESUL I  
AB076453

ABQ76453  
ID ABQ76453 standard: cDNA: 1631 BP.

AC ABO76453:

21-NOV-2002 (first entry)

DE S. cerevisiae BAX-associated cDNA fragment SEQ ID 331.

Bax; Bax-resistance; cytostatic; fungicide; immunosuppressive; virucide; vaxotropic; vaccine; gene therapy; proliferative disorder; cancer; apoptosis; fungal; yeast; infection; autoimmune disease; ischaemia; neurodegeneration; cell death; ss.

OS *Saccharomyces cerevisiae*.

PN WO200264766-A2.

22-AUG-2002.

PF 21-DEC-2001; 2001WO-EP15398.

PR 22-DEC-2000; 2000EP-0870318.

PR 04-JAN-2001; 2001EP-0870002.

PR 09-JAN-2001; 2001EP-0870003.

PA (JANC) JANSSEN PHARM NV.

PI Contreras RH, Eberhardt I, Luyten WHML, Reekmans RJ;

DR WPI; 2002-667002/71.

DR P-PSDB; ABG93187.

PT New isolated nucleic acid representing a synthetic Bax gene, useful as  
PT medicament for treating, preventing and/or alleviating yeast or fungal  
PT infections or proliferative disorders, or for preventing apoptosis in  
PT certain diseases -  
PT

PS Claim 36; Figure 1; 344pp; English.

This invention describes a novel nucleic acid representing a synthetic Bax gene. The Bax gene of the invention is useful for identifying Bax-resistant yeast or fungi, identifying, or obtaining and identifying *Candida* spp. sequences that are differentially expressed in a pathway eventually leading to programmed cell death or identifying inhibitors or inhibitor sequences of Bax-induced cell death. The products of the invention have cytostatic, fungicidal, immunosuppressive, virucide and vasotropic activity and can be used in vaccines or for gene therapy. The isolated nucleic acids, polypeptides, pharmaceutical compositions, antisense molecules and antibodies are useful as medicaments or in preparing a medicament for treating, preventing and/or alleviating diseases associated with yeast or fungi or proliferative disorders, such as cancer, or for preventing apoptosis in certain diseases. The compound or polypeptides, or the genetically modified organism are useful for preparing a medicament for modifying the endogenous flora of humans and other mammals. The vaccine is useful for immunising against yeast or fungal infections. Apoptosis-related diseases include autoimmune disease, ischaemia, diseases related with viral infections or neurodegenerations.



PI Contreras RH, Eberhardt I, Luyten WHML, Reekmans RJ;  
XX WPI: 2002-667002/71.  
DR P-PSDB; ABG93300.  
XX  
XX New isolated nucleic acid representing a synthetic Bax gene, useful as  
PT medicament for treating, preventing and/or alleviating yeast or fungal  
PT infections or proliferative disorders, or for preventing apoptosis in  
PT certain diseases -  
XX  
XX Claim 36; Figure 2; 344pp; English.  
XX  
XX This invention describes a novel nucleic acid representing a synthetic  
CC Bax gene. The Bax gene of the invention is useful for identifying  
CC Bax-resistant yeast or fungi, identifying, or obtaining and identifying  
CC Candida spp. sequences that are differentially expressed in a pathway  
CC eventually leading to programmed cell death or identifying inhibitors or  
CC inhibitor sequences of Bax-induced cell death. The products of the  
CC invention have cytostatic, fungicide; immunosuppressive, virucide and  
CC vasotropic activity and can be used in vaccines or for gene therapy. The  
CC isolated nucleic acids, polypeptides, pharmaceutical compositions,  
CC antisense molecules and antibodies are useful as medicaments or in  
CC preparing a medicament for treating, preventing and/or alleviating  
CC diseases associated with yeast or fungi or proliferative disorders, such  
CC as cancer, or for preventing apoptosis in certain diseases. The compounds  
CC or polypeptides, or the genetically modified organism are useful for  
CC preparing a medicament for modifying the endogenous flora of humans and  
CC other mammals. The vaccine is useful for immunising against yeast or  
CC fungal infections. Apoptosis-related diseases include autoimmune disease,  
CC ischaemia, diseases related with viral infections or neurodegenerations.  
CC This sequence represents a polynucleotide associated with the Bax gene  
CC described in the disclosure of the invention.  
XX  
XX Sequence 1086 BP; 359 A; 174 C; 223 G; 330 T; 0 other;  
SQ

Query Match 6.4%; Score 71; DB 24; Length 1086;  
Best Local Similarity 49.0%; Pred. No. 4.2e-09;  
Matches 324; Conservative 0; Mismatches 310; Indels 27; Gaps 4;  
35 TTGAGGTGCAAGCTGTTGTTAAACAGATGCTCAGTTCAGATTAAGAGGGTA 94  
Db 29 TCGAACCTTATCATATTAGCGGAATTAAGATATTCCTCAACAAAAATAAGAAATG 88  
95 CAGCCTTGGTGAAGTTGAGCTGCTGCTGTACCCCACTGATTGGAACATATTCCTT 154  
Db 89 AATATTAAACAAGCAGTAGCTTTTGCATTAACCACTGATTGGAACACATGTTT 148  
95 ATAAGATTGTTCCAGAAAGTTCAATTCAGGATGTGACATTCGCTGGTACATGTTCAAC 214  
Db 149 ATCAATTGGGCAGCCAGGTGATGCTGTTGCTGCGATGTTAGTGGGATCATTTGAAGA 208  
95 TTGACCAATGCTAGTACTGACTTGAAGTTGAGATACCGTTTCGGTTTGTTCACG 274  
Db 209 TGGGTTCTCAAG----TAACTGGGTTTGCAAAAGGTGACACTGTAAGTGTCTTTATAC 265  
95 GTGCTTCCCAACAGATCTTAAATGCTGATTTGCTGAATATGCCAGGTTTATCCAC 334  
Db 266 GTA-----ATGATCACTCGCACTGGAGCTTTTGAGAAATATGAGCTGTGATCCTG 319  
95 CTTTGTGTTTCAAGAGTAACCTTAACCTCACTCACTGCTGATGAAATTTCTGAAGGCCCTG 394  
Db 320 CTACTTCGATAAAGTACAATAAGAATTTTGAACATTTTGACTAATTTACAGATATCTGAAA 379  
95 TGAAGAACTTCGATCTGCTGCATCATTCGCAGTTTCGTTGACACTGCTGGTTAGTT 454  
Db 380 TCCACTCATTTGAAGGGGACGACGATTAATTTAGGTTTGGTTACCGTTGGGCTTTTCA 439  
95 TGTGTCATCACTTGGGCTCAAAAATGAATGGAATGGCAACCATCTACCCCGCAACATCTATC 514  
Db 440 TTCTCTATTACTTACGAATTTGACAAACAAAGCAACC-----TGGGGATA 484  
95 CATTTATTGTTGGGTTGGTCTACAGCAGTGGGTCAACAACTAATCAAGTTGCCAAAC 574

485 GTATTTTGTATTTGGGGAGGAGCAACTGCAACTGGAGTTCTAGCCATTCAGTTGCCAAAC 544  
Qy 575 ATATCAATGCTTATACTAAGATTGTAACTGTTCTTCTAAAGAGCATGAAAGCTTTTAA 634  
Db 545 TAGTGTA---TAATCTCAAGTAATCACCACAGCATCACCCAAAACACACCCCTTTGA 601  
Qy 635 AGTCTTATGCTGTGATGATGCTTTTGACTATCATGATGCGAGCGTTATTTGACGATCA 694  
Db 602 AACAAATTAGGGGAGATTAATGTTTTCGATTATGAGACGCTGATGTTGTCATAAATA 661  
Qy 695 A 695  
Db 662 A 662

RESULT 4  
ABT11258  
ID ABT11258 standard; DNA; 1330 BP.  
XX  
XX AC ABT11258;  
XX  
XX DT 10-DEC-2002 (first entry)  
XX  
XX Yeast selected interacting domain coding sequence SEQ ID NO: 15.  
XX  
XX Yeast; protein-protein interaction; Selected Interacting Domain;  
KW SID (RTM); secretion yield; cancer; neurodegenerative disease; fungicide;  
KW cytostatic; neuroprotective; gene; ds.  
XX  
XX Saccharomyces cerevisiae.  
XX  
XX WO200266504-A2.  
XX  
XX PD 29-AUG-2002.  
XX  
XX 14-FEB-2002; 2002WO-EP02299.  
XX  
XX PR 16-FEB-2001; 2001US-269266P.  
XX  
XX (HYBR-) HYBRIGENICS.  
XX  
XX PI Legrain P;  
XX  
XX WPI: 2002-674913/72.  
DR P-PSDB; ABJ10941.  
XX  
XX New protein-protein complexes of Saccharomyces cerevisiae, useful in  
PT drug screening or development, for developing yeast strains with better  
PT secretion yield of protein, or in gene therapy (e.g. to treat Candida  
PT infection or cancer)  
XX  
XX Claim 7; Page 41; 357pp; English.  
XX  
XX The present invention relates to complexes between Saccharomyces  
CC cerevisiae Selected Interacting Domain (SID (RTM)) proteins and coding  
CC sequences. The protein complexes of S. cerevisiae are useful in drug  
CC development, in screening drugs or agents that modulate the interaction  
CC of proteins, for developing yeast strains with better secretion yield of  
CC protein, and in gene therapy. The protein complexes, polypeptides and  
CC polynucleotides are useful for preventing or treating Candida infection,  
CC cancer or neurodegenerative diseases in humans or animals. The present  
CC sequence is a coding sequence of the invention.  
XX  
XX Sequence 1330 BP; 420 A; 257 C; 207 G; 446 T; 0 other;  
SQ

Query Match 5.8%; Score 64.4; DB 24; Length 1330;  
Best Local Similarity 56.0%; Pred. No. 2.9e-07;  
Matches 122; Conservative 0; Mismatches 96; Indels 0; Gaps 0;  
895 CAAGAAATTCATTGGGTGCAACAGATTTCTGCTAGTCCAGATATCATGAGCCACA 954  
Db 1 CATGAAGTACCATTGGAACACATTACTTTACCAGCGCGCTCAGAAAGCTGCA 60

Qy	955	GTTAAATTCGGTTAAGTTTATAAATCCACACTTAAACAACGGTGATATCCCATATGTAAT	1014
Db	61	ATAAAATTTATCAAATTCAATCAATCCTTAAAGAATTAATGATCGACAAATTCGCCATATTCCA	120
Qy	1015	ATTTAAAGTTTTCAGCAACGGCTTAGATGATGTCTCCAGCTCTCACTGAAGGTATATAAGAA	1074
Db	121	GTAAAGGCTCATAGAACGGCTTTGTGATGTTCTCATATCTCTAAAGAAGACATCAAAATAT	180
Qy	1075	GGTAAAAACAAAAATGTTAAAGTATGTTGCCAGGTTATA	1112
Db	181	GSTAAGAAGCTCTCGTGAAACAACTCGTTGCCGTATTAA	218
 RESULT 5 AAV74349/c			
XX	AAV74349	standard; DNA; 23439 BP.	
XX	AAV74349;		
DT	16-MAR-1999	(first entry)	
XX	Staphylococcus aureus	contig SEQ ID #38.	
KW	Computer readable medium; vaccine; S.aureus infection; immunodetection;		
KW	cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;		
KW	skin infection; surgical wound infection; scalded skin syndrome;		
KW	toxic shock syndrome; ds.		
XX			
OS	Staphylococcus aureus.		
XX			
FH	Key	Location/Qualifiers	
FT	misc_feature	481..540	
FT		/tag= a	
FT		/note= "these bases represent a line of missing text in	
FT		the sequence listing in the specification. They	
FT		are included to maintain the nucleotide numbering	
FT		given in the specification for this DNA sequence"	
FT	misc_feature	2281..2340	
FT		/tag= b	
FT		/note= "these bases represent a line of missing text in	
FT		the sequence listing in the specification. They	
FT		are included to maintain the nucleotide numbering	
FT		given in the specification for this DNA sequence"	
FT	misc_feature	4081..4140	
FT		/tag= c	
FT		/note= "these bases represent a line of missing text in	
FT		the sequence listing in the specification. They	
FT		are included to maintain the nucleotide numbering	
FT		given in the specification for this DNA sequence"	
FT	misc_feature	5881..5940	
FT		/tag= d	
FT		/note= "these bases represent a line of missing text in	
FT		the sequence listing in the specification. They	
FT		are included to maintain the nucleotide numbering	
FT		given in the specification for this DNA sequence"	
FT	misc_feature	7681..7740	
FT		/tag= e	
FT		/note= "these bases represent a line of missing text in	
FT		the sequence listing in the specification. They	
FT		are included to maintain the nucleotide numbering	
FT		given in the specification for this DNA sequence"	
FT	misc_feature	9481..9540	
FT		/tag= f	
FT		/note= "these bases represent a line of missing text in	
FT		the sequence listing in the specification. They	
FT		are included to maintain the nucleotide numbering	
FT		given in the specification for this DNA sequence"	
FT	misc_feature	11281..11340	
FT		/tag= g	
FT		/note= "these bases represent a line of missing text in	
FT		the sequence listing in the specification. They	
FT		are included to maintain the nucleotide numbering	
FT		given in the specification for this DNA sequence"	
FT			

CC (and their fragments) are useful as primers or probes for isolating  
CC homologues of any of the *S. aureus* DNA sequences contained on the  
CC computer readable medium.  
XX  
SQ Sequence 23439 BP; 7768 A; 3506 C; 4455 G; 6921 T; 789 other;  
Query Match 4.1%; Score 45.4; DB 18; Length 23439;  
Best Local Similarity 45.4%; Pred. No. 0.14;  
Matches 163; Conservative 0; Mismatches 196; Indels 0; Gaps 0;  
Qy 504 ACATACATCATCATATGATTTGGGGTGGTGTACAGCAGTGGTCAACAACTAATCCA 563  
Db 16078 ACAACGTTTTTAACTACGTTTATGACGTCGACAGGTAATGATGATTAATGATTA 16019  
Qy 564 AGTTGCCAAACATCATCATGCTTATACATGATTTGTAAGTGTCTTCTTAAAGCATGA 623  
Db 16018 CATTCGCTGATTAATTAATCAAAATATAATTTGCACAGATGAATTAACACACTTC 15959  
Qy 624 AAAGCTTTTAAAGTCTTATGCTGTGATGATGCTTTTGAATCATCATGATGCGGCTTAT 683  
Db 15958 ACCTGTTTTCATTCGGCATTTGATGGCATTTATGATGTTCCCGTTGAGGAACGAACA 15899  
Qy 684 TGACAGATCAAAATCGAAGTATCCAAACCTGCAACATGTTTATGACGCTGTGGGAACGA 743  
Db 15898 TATCATGAATCATGTTCCACATTTCAACATTTGAAACGGGTGAATAAAAAACGACATGTT 15839  
Qy 744 AGATAGTATCCCGAGGCTTAAAGTCACAGCAGATAGTCTACCTGCCACATTTAGA 803  
Db 15838 TGATCGTAGACCTAATGACGAAGCAATCACTATTTATCGTAAAGTGTGTTGATTTCTTAA 15779  
Qy 804 AGTGGTTCATGACCATTTGAAAGCATTCCTGGAAGAAATCAGAAAGATAATGTTAAAA 862  
Db 15778 CGCTATACATGATGATTAATATATCATTAATTTTAAAAAACTCAATTTATCATCATGATA 15720

## RESULT 6

AAA70202  
ID AAA70202 standard; DNA; 3501 BP.  
AC AAA70202;  
XX  
XX 07-NOV-2000 (first entry)  
XX Plasmodium falciparum chromosome 2 related DNA sequence SEQ ID NO:335.  
XX Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;  
KW antimalarial; malaria; protozoa; infection; insecticide; ds.  
XX  
XX Plasmodium falciparum.  
OS  
XX W0200025728-A2.  
XX  
XX 11-MAY-2000.  
XX  
XX 05-NOV-1999; 99WO-US26796.  
XX  
XX 05-NOV-1998; 98US-0107131.  
XX  
XX (HOFF/) HOFFMAN S.  
PA (CARU/) CARUCCI D.  
PA (GARD/) GARDNER M.  
PA (VENT/) VENTER J C.  
XX

PI Hoffman S, Carucci D, Gardner M, Venter JC;  
XX

DR WPI; 2000-365347/31.  
XX

XX Proteins encoded by chromosome 2 of the human malaria parasite,  
PT Plasmodium falciparum, useful as antimalarial vaccines and in the  
PT diagnosis of *P. falciparum* infection -  
XX

PS Disclosure; Page 528-529; 577pp; English.  
XX

CC The present invention describes proteins and their fragments (I) encoded  
CC by chromosome 2 of the human malaria parasite, *Plasmodium falciparum*.  
CC Also described are: (I) nucleotide sequences (II) encoding (I); and (2)  
CC vaccines against *P. falciparum* infection comprising (I) or (II).  
CC (I) and (II) are useful for the development of vaccines against  
CC *P. falciparum* infection. (I) and polyclonal antisera or a monoclonal  
CC antibody raised to immunogens comprising the sequences of (I), are  
CC useful in the detection of infection with *P. falciparum*. Furthermore,  
CC (I) especially when they are rifins or secreted or membrane proteins)  
CC can aid the identification of drugs to treat or prevent *P. falciparum*  
CC infection, or they can be used to identify drug resistance in  
CC *P. falciparum*. Sequencing of the *Plasmodium* chromosome 2 and the  
CC subsequent identification of proteins encoded by it will help to expand  
CC our understanding of parasite biology, a process hampered by the  
CC complexity of the parasitic lifecycle, and provide new targets for  
CC vaccine and drug development. Parasite resistance to drugs and mosquito  
CC resistance to insecticides have led to a resurgence of malaria in many  
CC parts of the world, and there is a pressing need for vaccines and new  
CC drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide  
CC and protein sequences given in the present invention, but which are not  
CC specifically mentioned within the specification.  
XX

SQ Sequence 3501 BP; 1699 A; 327 C; 468 G; 1007 T; 0 other;

Query Match 3.7%; Score 40.8; DB 21; Length 3501;  
Best Local Similarity 46.2%; Pred. No. 1.2;  
Matches 209; Conservative 0; Mismatches 237; Indels 6; Gaps 2;  
Qy 650 ATGATGCTTTTGAATCATGATGCGAGCGGTTATTGAGCAGATCAAAATCGAAGTATCCAA 709  
Db 194 ATGATGAATGTTGTTAAGACGCGCTGAAAGTGAATGAGGAGATCCCAACAAAGAAAACGA 253  
Qy 710 ACCCTGCAACATGTTATTGACGCTGTGGGAAGCGAAGATAGTATCCCGAGGCTATAAAG 769  
Db 254 ACCATGATTAATATAAACAATAATGATGTAATGATGATTAATATATATATATATATATAT 310  
Qy 770 TCACAGCAGATAGTCTACCTGCCACATTTATTAGAAGTGTTCCTCAATGACCATGAAAGCA 829  
Db 311 TAAATGATGATTAATATAGCAATAATCATATATAGATGATGATTAATATAAACAATAATCAT 370  
Qy 830 TTCTTGAAGAAATCAGAAAAGATTAATGTTAAATTTGATATTAATTTGTTGTATCGTGCAT 889  
Db 371 TAAATGATGATTAATATAAACAATAATGATGTTAAATGATGATTAATATAAATAATCAAT 430  
Qy 890 CTGGTCAAGAAATCTTATTGGGTGCAACAGATTTCTCTGCTAGTCAGATATCATGAAG 949  
Db 431 TAAATGATGATTAATATAAATAATTAATCATATATAATGATGATTAATATAAATAATCA 490  
Qy 950 CCACAGTTAAATTCGTTAAAGTTTATAAATCCACACCTTAACACGGTGATATCCATCATA 1009  
Db 491 TAAATGATGATTAATATAAACAATAATGATTAATGATGATTAATATAAATAATTAATAT 550  
Qy 1010 TGAATATTAAAGTTTTTCAGCAACGGCTTAGATG---ATGTCCAGCTCTCAGTGAAGGTA 1066  
Db 551 GTAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 610  
Qy 1067 TAAAGCAAGGTAAACAAACAAATGTTAAGTAT 1098  
Db 611 AAAAGCAAGGTAAACAAACAAATGTTAAGTAT 642

## RESULT 7

ABA49878/c  
ID ABA49878 standard; DNA; 439 BP.  
XX  
XX ABA49878;  
XX  
XX 01-FEB-2002 (first entry)  
XX Human breast cell single exon nucleic acid probe #8573.  
DE Human; microarray; single exon probe; gene expression; breast;  
KW disease; cancer; ss.  
XX











CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 439 BP; 141 A; 169 C; 41 G; 88 T; 0 other;  
Query Match 3.6%; Score 40.2; DB 22; Length 439;  
Best Local Similarity 46.8%; Pred. No. 0.74;  
Matches 126; Conservative 0; Mismatches 143; Indels 0; Gaps 0;  
QY 36 TGAAGGTGACAAAGCTGTTGTTAAACAGATGCTCAGTCCAGATTAAGGAGGCTAC 95  
DB 385 TGATGGTGGTATGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATG 326  
QY 96 AGCCTTGGTGAAGGCTGTTGCTGTTAAACAGATGCTCAGTCCAGATTAAGGAGGCTAC 155  
DB 325 TGCCGATGGTATGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATG 266  
QY 156 TAAGATTGGTCCAGAGGTTCAATCTAGGATGATGATGATGATGATGATGATGATGAT 215  
DB 265 TGCTGGTGGTATGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATG 206  
QY 216 TGGACCAAACTAGTACTGCTGTTGAAAGGTTGGAGATACCGGTTTCGTTTTCACGG 275  
DB 205 TCCTGCTGATGCTGTTGATGCTGTTGATGATGATGATGATGATGATGATGATGATG 146  
QY 276 TGCTTCCCAACACAGATCCTAAAAATGGTG 304  
DB 145 TGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 117  
RESULT 15  
ABS41558/c  
ID ABS41558 standard; DNA; 439 BP.  
XX  
AC ABS41558;  
XX  
XX 25-FEB-2003 (first entry)  
XX  
XX Human liver single exon probe, SEQ ID No 16548.  
XX  
XX Human; single exon nucleic acid probe; liver; cirrhosis;  
KW hyperlipoproteinaemia; hyperlipidaemia; hypercholesterolaemia;  
KW coronary heart disease; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO200157273-A2.  
XX  
XX 09-AUG-2001.  
XX  
XX 30-JAN-2001; 2001WO-US00664.  
XX  
XX 04-FEB-2000; 2000US-0180312.  
XX  
XX 26-MAY-2000; 2000US-0207456.  
XX  
XX 30-JUN-2000; 2000US-0608408.  
XX  
XX 03-AUG-2000; 2000US-0632366.  
XX  
XX 21-SEP-2000; 2000US-0234687.  
XX  
XX 27-SEP-2000; 2000US-0236359.  
XX  
XX 04-OCT-2000; 2000GB-0024253.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
XX WPI; 2001-488898/53.  
XX  
XX Human genome-derived single exon nucleic acid probes useful for  
PT analysing gene expression in human adult liver -  
XX  
XX Claim 4; SEQ ID No 16548; 658pp; English.  
XX  
XX The invention relates to a single exon nucleic acid probe (SENP) (I) for  
CC measuring human gene expression in a sample derived from human adult

CC liver, comprising one of 13109 defined nucleotide sequences given in the  
CC specification (or complements/fragments). The probe hybridises at high  
CC stringency to a nucleic acid molecule expressed in the human adult  
CC liver. (I) may be used for predicting, measuring and displaying gene  
CC expression in samples derived from human adult liver. The genes  
CC identified may be involved in genetic liver diseases such as cirrhosis,  
CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which  
CC is associated with coronary heart disease. ABS25011-ABS51005 represent  
CC human liver single exon nucleic acid probes of the invention.  
CC Note: The sequence information for this patent does not appear in the  
CC printed specification but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 439 BP; 141 A; 169 C; 41 G; 88 T; 0 other;  
Query Match 3.6%; Score 40.2; DB 23; Length 439;  
Best Local Similarity 46.8%; Pred. No. 0.74;  
Matches 126; Conservative 0; Mismatches 143; Indels 0; Gaps 0;  
QY 36 TGAAGGTGACAAAGCTGTTGTTAAACAGATGCTCAGTCCAGATTAAGGAGGCTAC 95  
DB 385 TGATGGTGGTATGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATG 326  
QY 96 AGCCTTGGTGAAGGCTGTTGCTGTTAAACAGATGCTCAGTCCAGATTAAGGAGGCTAC 155  
DB 325 TGCCGATGGTATGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATG 266  
QY 156 TAAGATTGGTCCAGAGGTTCAATCTAGGATGATGATGATGATGATGATGATGATGAT 215  
DB 265 TGCTGGTGGTATGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATG 206  
QY 216 TGGACCAAACTAGTACTGCTGTTGAAAGGTTGGAGATACCGGTTTCGTTTTCACGG 275  
DB 205 TCCTGCTGATGCTGTTGATGCTGTTGATGATGATGATGATGATGATGATGATGATG 146  
QY 276 TGCTTCCCAACACAGATCCTAAAAATGGTG 304  
DB 145 TGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 117  
RESULT 16  
ABS15980/c  
ID ABS15980 standard; DNA; 439 BP.  
XX  
AC ABS15980;  
XX  
XX 19-AUG-2002 (first entry)  
XX  
XX Human genome-derived single exon probe ORF from lung SEQ ID No 15971.  
XX  
XX Human; ds; single exon probe; asthma; lung cancer; COPD; IUD;  
KW chronic obstructive pulmonary disease; interstitial lung disease;  
KW familial idiopathic pulmonary fibrosis; neurofibromatosis;  
KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;  
KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;  
KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;  
KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;  
KW primary ciliary dyskinesia; pulmonary hypertension;  
KW hyaline membrane disease; open reading frame; ORF.  
XX  
XX Homo sapiens.  
XX  
XX WO200186003-A2.  
XX  
XX 15-NOV-2001.  
XX  
XX 30-JAN-2001; 2001WO-US00665.  
XX  
XX 04-FEB-2000; 2000US-180312P.  
XX  
XX 26-MAY-2000; 2000US-207456P.  
XX  
XX 30-JUN-2000; 2000US-0608408.  
XX  
XX 03-AUG-2000; 2000US-0632366.  
XX  
XX 21-SEP-2000; 2000US-0234687P.

PR 27-SEP-2000; 2000US-236359P.  
PR 04-OCT-2000; 2000GB-0024263.  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2002-114183/15.  
XX  
XX Spatially-addressable set of single exon nucleic acid probes, used to  
PT measure gene expression in human lung samples -  
XX  
XX Claim 4; SEQ ID No 15971; 634pp; English.  
XX  
XX The invention relates to a spatially-addressable set of single exon  
CC nucleic acid probes for measuring gene expression in a sample derived  
CC from human lung comprising single exon nucleic acid probes having one of  
CC 12614 nucleic acid sequences mentioned in the specification, or their  
CC complements or the 12387 open reading frames derived from the 12614  
CC probes. Also included are a microarray comprising the novel set of  
CC probes; the novel set of probes which hybridize at high stringency to a  
CC nucleic acid expressed in the human lung; measuring gene expression in a  
CC sample derived from human lung, comprising (a) contacting the array with  
CC a collection of detectably labeled nucleic acids derived from human lung  
CC mRNA, and (b) measuring the label detectably bound to each probe of  
CC the array; identifying exons in a eukaryotic genome, comprising  
CC (a) algorithmically predicting at least one exon from genomic sequences  
CC of the eukaryote; and (b) detecting specific hybridisation of detectably  
CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,  
CC having a fragment identical to the predicted exon, the probe is included  
CC in the above mentioned microarray; assigning exons to a single gene,  
CC comprising (a) identifying exons from genomic sequence by the method  
CC above and (b) measuring the expression of each of the exons in several  
CC tissues and/or cell types using hybridisation to a single exon  
CC microarrays having a probe with the exon, where a common pattern of  
CC expression of the exons in the tissues and/or cell types indicates that  
CC the exons should be assigned to a single gene; a peptide comprising one  
CC of 12011 sequences, mentioned in the specification, or encoded by the  
CC probes/open reading frames (ORF). The probes are used for gene  
CC expression analysis, and for identifying exons in a gene, particularly  
CC using human lung derived mRNA and for the study of lung diseases  
CC such as asthma, lung cancer, chronic obstructive pulmonary disease  
CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary  
CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,  
CC Niemann-Pick disease, Hermannsky-Pudlak syndrome, sarcoidosis, pulmonary  
CC haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis,  
CC pulmonary alveolar proteinosis, Karagener syndrome, fibrocystic  
CC pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension  
CC and hyaline membrane disease. The present sequence is a single exon  
CC probe open reading frame of the invention.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WPIPO at  
XX ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 439 BP; 141 A; 169 C; 41 G; 88 T; 0 other;  
  
Query Match 3.6%; Score 40.2; DB 24; Length 439;  
Best Local Similarity 46.8%; Pred. No. 0.74;  
Matches 126; Conservative 0; Mismatches 143; Indels 0; Gaps 0;  
  
QY 36 TGAAGGTGACAAAGCTGTTGTTAAACAGATGCTCAGTTCACAGATTAAAGGAGGTAC 95  
Db |||||  
385 TGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 326  
QY 96 AGCCTTGGTGAAGTTGAGGCTGTTGCTGGTGAACCCAACTGATTGGAAGCATATTGCTTA 155  
Db |||||  
325 TGCCGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 266  
QY 156 TAAGATTGGTCCAGAAAGTTCAATTCTAGGATGGACATTTGCTGGTGAAGTTGCAAACT 215  
Db |||||  
265 TGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 206

QY 216 TGGACCAAAATGCTAGTACTGACTTGAAGGTGGAGATACCGGTTTCGGTTTGTTCACGG 275  
Db |||||  
205 TCCTGCTGATGGTTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 146  
QY 276 TGGTTCCCAACAGATCCTTAAATAATGGTG 304  
Db |||||  
145 TGATGATGATGGTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 117

RESULT 17  
AAD44410/c  
ID AAD44410 standard; DNA; 1236 BP.  
XX  
XX AAD44410;  
XX  
XX 13-DEC-2002 (first entry)  
XX Human huntington (htQ103) protein encoding DNA.  
XX  
XX Human; protein misfolding; Alzheimer's disease; AD; Parkinson's disease;  
XX PD; Familial amyloid polyneuropathy; tauopathy; frontotemporal dementia;  
XX Pick disease; lobar atrophy; trinucleotide disease; fragile-X syndrome;  
XX Huntington's disease; spinocerebellar ataxia; SCA; myotonic dystrophy;  
XX dentatorubral pallidoluysian atrophy; DRPLA; Creutzfeldt-Jacob disease;  
XX CJD; prion disease; Gerstmann-Strausler-Scheinker disease; GSS; PFI;  
XX fatal familia insomnia; mad cow disease; scrapie; kuru; anticonvulsant;  
XX neurotropic; neuroprotective; cerebroprotective; htQ103 protein; gene; ds.  
XX  
XX Homo sapiens.  
XX  
XX Key Location/Qualifiers  
XX CDS 1..513  
XX /tag= a  
XX /product= "Human htQ103 protein"  
XX /note= "No stop codon"  
XX /partial  
XX  
XX WO200265136-A2.  
XX  
XX 22-AUG-2002.  
XX  
XX 15-FEB-2002; 2002WO-US04632.  
XX  
XX 15-FEB-2001; 2001US-269157P.  
XX (UYCH-) UNIV CHICAGO.  
XX  
XX Lindquist S, Krobitsch S, Outeiro T;  
XX  
XX WPI; 2002-667026/71.  
XX P-PSDB; AAE26650.  
XX  
XX Screening for therapeutic agents for protein misfolding disease, by  
XX contacting a yeast cell with compound, that expresses misfolded disease  
XX protein, and with a toxicity inducing agent, and evaluating cell for  
XX viability -  
XX  
XX Disclosure; Page 87-88; 93pp; English.  
XX  
XX The present invention relates to novel screening methods for identifying  
XX therapeutic agents for diseases associated with protein misfolding. The  
XX method involves contacting a yeast cell with a candidate compound, where  
XX the yeast cell expresses a polypeptide comprising a misfolded disease  
XX protein, contacting the yeast cell with a toxicity inducing agent and  
XX evaluating the yeast cell for viability, where the viability indicates  
XX the candidate compound is a candidate therapeutic agent. The method is  
XX useful to screen for therapeutic agents for diseases associated with  
XX protein misfolding such as Alzheimer's disease (AD), Parkinson's disease  
XX (PD), Familial amyloid polyneuropathy, tauopathies (e.g. Pick disease,  
XX lobar atrophy, frontotemporal dementia) or trinucleotide diseases (e.g.  
XX Huntington's disease, spinocerebellar ataxia (SCA), fragile-X syndrome,  
XX myotonic dystrophy, dentatorubral pallidoluysian atrophy (DRPLA) and  
XX prion diseases (e.g. Creutzfeldt-Jacob disease (CJD), fatal familia



PR 31-JAN-2000; 2000US-0179065.  
 PR 04-FEB-2000; 2000US-0180628.  
 PR 24-FEB-2000; 2000US-0184664.  
 PR 02-MAR-2000; 2000US-0186350.  
 PR 16-MAR-2000; 2000US-0189874.  
 PR 17-MAR-2000; 2000US-0190076.  
 PR 18-APR-2000; 2000US-0198123.  
 PR 19-MAY-2000; 2000US-0205515.  
 PR 07-JUN-2000; 2000US-0209467.  
 PR 28-JUN-2000; 2000US-0214886.  
 PR 30-JUN-2000; 2000US-0215135.  
 PR 07-JUL-2000; 2000US-0216647.  
 PR 07-JUL-2000; 2000US-0216880.  
 PR 11-JUL-2000; 2000US-0217487.  
 PR 11-JUL-2000; 2000US-0217496.  
 PR 14-JUL-2000; 2000US-0218290.  
 PR 26-JUL-2000; 2000US-0220963.  
 PR 28-JUL-2000; 2000US-0220964.  
 PR 14-AUG-2000; 2000US-0224518.  
 PR 14-AUG-2000; 2000US-0224519.  
 PR 14-AUG-2000; 2000US-0225213.  
 PR 14-AUG-2000; 2000US-0225214.  
 PR 14-AUG-2000; 2000US-0225266.  
 PR 14-AUG-2000; 2000US-0225267.  
 PR 14-AUG-2000; 2000US-0225268.  
 PR 14-AUG-2000; 2000US-0225270.  
 PR 14-AUG-2000; 2000US-0225447.  
 PR 14-AUG-2000; 2000US-0225757.  
 PR 14-AUG-2000; 2000US-0225758.  
 PR 14-AUG-2000; 2000US-0225759.  
 PR 18-AUG-2000; 2000US-0226279.  
 PR 22-AUG-2000; 2000US-0226681.  
 PR 22-AUG-2000; 2000US-0226686.  
 PR 22-AUG-2000; 2000US-0226982.  
 PR 23-AUG-2000; 2000US-0227009.  
 PR 30-AUG-2000; 2000US-0228924.  
 PR 01-SEP-2000; 2000US-0229287.  
 PR 01-SEP-2000; 2000US-0229343.  
 PR 01-SEP-2000; 2000US-0229344.  
 PR 01-SEP-2000; 2000US-0229345.  
 PR 03-SEP-2000; 2000US-0229509.  
 PR 05-SEP-2000; 2000US-0229513.  
 PR 06-SEP-2000; 2000US-0230437.  
 PR 06-SEP-2000; 2000US-0230438.  
 PR 08-SEP-2000; 2000US-0231242.  
 PR 08-SEP-2000; 2000US-0231243.  
 PR 08-SEP-2000; 2000US-0231244.  
 PR 08-SEP-2000; 2000US-0231413.  
 PR 08-SEP-2000; 2000US-0231414.  
 PR 08-SEP-2000; 2000US-0232080.  
 PR 08-SEP-2000; 2000US-0232081.  
 PR 12-SEP-2000; 2000US-0231968.  
 PR 14-SEP-2000; 2000US-0232397.  
 PR 14-SEP-2000; 2000US-0232398.  
 PR 14-SEP-2000; 2000US-0232399.  
 PR 14-SEP-2000; 2000US-0232400.  
 PR 14-SEP-2000; 2000US-0232401.  
 PR 14-SEP-2000; 2000US-0233063.  
 PR 14-SEP-2000; 2000US-0233064.  
 PR 14-SEP-2000; 2000US-0233065.  
 PR 21-SEP-2000; 2000US-0234223.  
 PR 21-SEP-2000; 2000US-0234274.  
 PR 25-SEP-2000; 2000US-0234997.  
 PR 25-SEP-2000; 2000US-0234998.  
 PR 26-SEP-2000; 2000US-0235484.  
 PR 27-SEP-2000; 2000US-0235834.  
 PR 27-SEP-2000; 2000US-0235836.  
 PR 29-SEP-2000; 2000US-0236327.  
 PR 29-SEP-2000; 2000US-0236367.  
 PR 29-SEP-2000; 2000US-0236368.  
 PR 29-SEP-2000; 2000US-0236369.  
 PR 29-SEP-2000; 2000US-0236370.  
 PR 02-OCT-2000; 2000US-0236802.

PR 02-OCT-2000; 2000US-0237037.  
 PR 02-OCT-2000; 2000US-0237038.  
 PR 02-OCT-2000; 2000US-0237039.  
 PR 02-OCT-2000; 2000US-0237040.  
 PR 13-OCT-2000; 2000US-0239935.  
 PR 13-OCT-2000; 2000US-0239937.  
 PR 20-OCT-2000; 2000US-0240960.  
 PR 20-OCT-2000; 2000US-0241221.  
 PR 20-OCT-2000; 2000US-0241785.  
 PR 20-OCT-2000; 2000US-0241786.  
 PR 20-OCT-2000; 2000US-0241787.  
 PR 20-OCT-2000; 2000US-0241808.  
 PR 20-OCT-2000; 2000US-0241809.  
 PR 20-OCT-2000; 2000US-0241826.  
 PR 01-NOV-2000; 2000US-0244617.  
 PR 08-NOV-2000; 2000US-0246474.  
 PR 08-NOV-2000; 2000US-0246475.  
 PR 08-NOV-2000; 2000US-0246476.  
 PR 08-NOV-2000; 2000US-0246477.  
 PR 08-NOV-2000; 2000US-0246478.  
 PR 08-NOV-2000; 2000US-0246523.  
 PR 08-NOV-2000; 2000US-0246524.  
 PR 08-NOV-2000; 2000US-0246525.  
 PR 08-NOV-2000; 2000US-0246526.  
 PR 08-NOV-2000; 2000US-0246527.  
 PR 08-NOV-2000; 2000US-0246528.  
 PR 08-NOV-2000; 2000US-0246532.  
 PR 08-NOV-2000; 2000US-0246609.  
 PR 08-NOV-2000; 2000US-0246610.  
 PR 08-NOV-2000; 2000US-0246611.  
 PR 08-NOV-2000; 2000US-0246613.  
 PR 17-NOV-2000; 2000US-0249207.  
 PR 17-NOV-2000; 2000US-0249208.  
 PR 17-NOV-2000; 2000US-0249209.  
 PR 17-NOV-2000; 2000US-0249210.  
 PR 17-NOV-2000; 2000US-0249211.  
 PR 17-NOV-2000; 2000US-0249212.  
 PR 17-NOV-2000; 2000US-0249213.  
 PR 17-NOV-2000; 2000US-0249214.  
 PR 17-NOV-2000; 2000US-0249215.  
 PR 17-NOV-2000; 2000US-0249216.  
 PR 17-NOV-2000; 2000US-0249217.  
 PR 17-NOV-2000; 2000US-0249218.  
 PR 17-NOV-2000; 2000US-0249244.  
 PR 17-NOV-2000; 2000US-0249245.  
 PR 17-NOV-2000; 2000US-0249264.  
 PR 17-NOV-2000; 2000US-0249265.  
 PR 17-NOV-2000; 2000US-0249297.  
 PR 17-NOV-2000; 2000US-0249299.  
 PR 01-DEC-2000; 2000US-0250160.  
 PR 01-DEC-2000; 2000US-0250391.  
 PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251988.  
 PR 05-DEC-2000; 2000US-0256719.  
 PR 06-DEC-2000; 2000US-0251479.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251868.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251989.  
 PR 11-DEC-2000; 2000US-0251990.  
 PR 05-JAN-2001; 2000US-0254097.  
 PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-483426/52.

Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
 useful for preventing, diagnosing and/or treating cancers and  
 metastasis -  
 PT

XX Disclosure; SEQ ID NO 22285; 3071pp + Sequence Listing; English.

XX AA54951 to AA64702 encode the human immune/haematopoietic antigen (I)

XX amino acid sequences given in AA82170 to AA91921. (I) have cytostatic

XX activity, and can be used in gene therapy and vaccine production. (I)

XX proteins and polynucleotides may be used in the prevention, diagnosis and

XX treatment of diseases associated with inappropriate (I) expression. For

XX example, they may be used to treat disorders associated with decreased

XX expression by rectifying mutations or deletions in a patient's genome

XX that affect the activity of (I) by expressing inactive proteins or to

XX supplement the patient's own production of (I). Additionally, (I)

XX polynucleotides may be used to produce the secreted (I), by inserting

XX the nucleic acids into a host cell and culturing the cell to express the

XX protein. (I) proteins and polynucleotides may be used to prevent,

XX diagnose and treat immune/haematopoietic-related diseases, especially

XX cancers and cancer metastases of haematopoietic-derived cells. AA64703

XX to AA87694 represent human immune/haematopoietic antigen genomic

XX sequences from the present invention. AA54942 to AA54950 and AA82169

XX represent sequences used in the exemplification of the present invention.

XX SQ Sequence 4027 BP; 1247 A; 875 C; 797 G; 1108 T; 0 other;

Query Match 3.6%; Score 39.6; DB 22; Length 4027;

Best Local Similarity 44.0%; Pred. No. 2.7;

Matches 168; Conservative 0; Mismatches 214; Indels 0; Gaps 0;

Qy 639 TTATGGTGCTGATGCTTTGACTATCATGATGCGAGCGGTTATTGAGCAGATCAATC 698

Dy 987 TTCAGCGATTCTCTTGCTCAGGCTCCCAAGTAGCTGGGATTACAGGTGTGCCACCA 1046

Qy 699 GAAGTATCCAACTCGCAACATGTTATTGACGCTGTGGAGCGAAGATAGTATCCCGA 758

Dy 1047 CACTCAGCTAATTTTGTATTTTAGTAGCGCAGGGTTTGGCATGTTACACGAGTGG 1106

Qy 759 GGCTATAAAGTCAGCAGATAGTCTACCTGCGCACATTTATAGAAGTGGTTCATGAC 818

Dy 1107 TCTCGAATCTCGACCTCAGGTGATCGCCCTGCGCTCCCAAAAGTGTGGGATTAC 1166

Qy 819 CATTTGAAGCATTCCTGAGGAATCAGAAAGATAAGTAAATGTTATGATATTACTTGT 878

Dy 1167 AGGCGTAGGCCACCAACCCAGCTTAATGATTAAGCAATGTTCTAATAAGAAATGGA 1226

Qy 879 GTATCGTGCTCTGCTCAAGAAATCTATTGGGTGCAACAGATTTCTGCTAGTCCAGA 938

Dy 1227 GAGAAACCATTTAAGGAAGTAATCAATTTAGTCAACGGGACTTCTGAGATTCACT 1286

Qy 939 ATATCATGAAGCCACAGTTAAATTCGTTAAGTTTATATAATCCACCTTAACAACGGTGA 998

Dy 1287 TACTTAAAGTGAATAATGTATACATTTTATTCTTCAAAACACAGAAACCTGGGTTA 1346

Qy 999 TATCCATCATATGATATTTAA 1020

Dy 1347 GATGGTTAAATGATGGTTAA 1368

RESULT 20

ABL55644/c

ID ABL55644 standard; DNA; 50000 BP.

XX AC ABL55644;

XX DT 01-JUL-2002 (first entry)

XX DE AmEPV genome fragment#2.

XX AmEPV; gene therapy; viral vector; chromosome mapping; gene mapping;

XX genetic deficiency disorder; da.

XX Amascta moorei entomopoxvirus.

XX WO200212526-A2.

XX

PD 14-FEB-2002.

XX 10-AUG-2001; 2001WO-US25287.

XX 10-AUG-2000; 2000US-224479P.

PR 14-SEP-2000; 2000US-0662254.

XX (UYFL ) UNIV FLORIDA.

XX Moyer RW, Li Y, Bawden AL;

PI WPI; 2002-227161/28.

XX Novel recombinant entomopox virus vector useful for delivering

PT polynucleotide encoding protein to vertebrate cell, comprises

PT polynucleotide encoding protein operably linked with heterologous

PT promoter sequence -

XX Disclosure; Page 150-175; 326pp; English.

XX The invention relates to a recombinant entomopox virus (EPV) vector,

CC comprising a polynucleotide encoding a protein operably linked with a

CC heterologous promoter sequence. The invention also concerns methods for

CC providing gene therapy for genetic deficiency disorders. Vectors of the

CC invention are useful for delivering a polynucleotide encoding a protein

CC to a vertebrate cell preferably a mammalian cell, such as a human cell.

CC The vector is introduced into the vertebrate cell by infection in a viral

CC particle, or by transfection, transduction, or injection either in vitro

CC or in vivo. The vector is useful for the delivery and expression of

CC biologically useful proteins in gene therapy protocols, and for

CC delivering large DNA segments for engineering of vertebrate cells.

CC Polynucleotides of the invention have applications in techniques such as

CC their use as insertion sites for foreign genes of interest, hybridisation

CC probes, for chromosome and gene mapping, in PCR technologies, and in the

CC production of sense or antisense nucleic acids. Vectors of the invention

CC provide for stable integration and expression of heterologous DNA in host

CC cells, and are adapted for accepting large heterologous polynucleotide

CC inserts which can be delivered in an infected or transformed cell and

CC expressed in a stable fraction. The current sequence represents a

CC fragment of the genome of the genus B entomopoxvirus from amascta moorei

CC (AmEPV).

XX SQ Sequence 50000 BP; 20514 A; 4505 C; 4614 G; 20367 T; 0 other;

Query Match 3.6%; Score 39.6; DB 24; Length 50000;

Best Local Similarity 49.5%; Pred. No. 7.5;

Matches 102; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

Qy 889 TCTGGTCAAGAAATTCATTGGGTGCAACAGATTTCTCTGCTAGTCCAGAAATATCATGAA 948

Dy 1619 TATTATGAAAAAATTTATTATATATAAAAAACCAATTTCTATGTATAGCAAAATATCATGAA 1560

Qy 949 GCCACAGTTAAATTCGTTAAGTTTATATAATCCACCTTAACAACGGTGTATCCATCAT 1008

Dy 1559 GAAAAAGATATTTATATCTATAGATATAAGATATAATCAATATGATGATATTCCTTAAAGAA 1500

Qy 1009 ATGAATATTAAGTTTTTTCAGCAACGGCTTAGATGATGCTCCAGCTCTCACTGAAGGTATA 1068

Dy 1499 AATATAAAAAAATTCITTTTATGATATATTTAATAAATACAGATATATTTTGAATAATTT 1440

Qy 1069 AAAGAAGGTAAAAACAAAAATGTTAA 1094

Dy 1439 AAAATTAAAAAATAATGTTGATTA 1414

RESULT 21

AAT91162

ID AAT91162 standard; cDNA; 772 BP.

XX AC AAT91162;

XX DT 27-MAR-1998 (first entry)

XX





AC AAT91136;  
 XX 27-MAR-1998 (first entry)  
 XX Strawberry fruit auxin-related gene.  
 DE Fruit ripening; strawberry; auxin-related gene; transgenic plant;  
 XX 88.  
 XX Fragaria x ananassa Duch cv. Brighton.  
 OS WO9727295-A1.  
 XX 31-JUL-1997.  
 XX 21-JAN-1997; 97WO-GB00178.  
 XX 09-SEP-1996; 96GB-0018742.  
 PR 23-JAN-1996; 96GB-0001330.  
 XX (HORT-) HORTICULTURE RES INT.  
 XX PA  
 XX Manning K;  
 XX WPI; 1997-393683/36.  
 DR  
 XX Modification of ripening of strawberry fruit - using vectors  
 PT comprising ripening sequences of e.g. O-methyl transferase, acyl  
 PT carrier protein, cellulase or cyteine protease  
 XX  
 XX Claim 2; Page 36-37; 72pp; English.  
 XX This cDNA clone encodes a strawberry auxin-related protein that is  
 CC implicated in fruit ripening-related processes. It was isolated  
 CC from a cDNA library of strawberry ripening genes by differential  
 CC screening using cDNA from fruit receptacle tissue at the ripe and  
 CC white stages of ripeness, and was identified by homology to  
 CC sequences in the EMBL database. A claimed vector for use in the  
 CC genetic transformation of strawberry cells comprises a promoter, a  
 CC regulation sequence and a transcription termination sequence. The  
 CC at least 10 bases) selected from 38 claimed strawberry fruit  
 CC ripening-related genes (see f91126-63) that is aligned in the  
 CC vector in sense or antisense direction. Phenotypically modified  
 CC strawberry plants can be generated that have one or more of the  
 CC following fruit characteristics: improved resistance to damage;  
 CC longer shelf-life; improved processing characteristics; improved  
 CC flavour and aroma; and modified colour.  
 XX  
 SQ Sequence 1151 BP; 308 A; 205 C; 292 G; 333 T; 13 other;  
 Query Match 3.5%; Score 39.4; DB 18; Length 1151;  
 Best Local Similarity 65.2%; Pred. No. 1.8; Gaps 0;  
 Matches 58; Conservative 0; Mismatches 31; Indels 0; Gaps 0;  
 QY 56 TTAACACAGATGCTCAGTTCCAGAAATTAAGGAGGTCACAGCCTTGGTGAAGTTGAG 115  
 DB 124 TTGATCCAGTGTGGCTGTTCTCTGAATTAAGAGGATCAGTCTCATCAGGTTGTTG 183  
 QY 116 CTGTTGCTGTTAACCACCACTGATTTGGAAG 144  
 DB 184 CTGCTTCTTTAACCACCACTGATTTGAAG 212  
 RESULT 24  
 ABN71527/c  
 ID ABN71527 standard; DNA; 2155561 BP.  
 XX  
 XX ABN71527;  
 XX  
 XX 02-JUL-2002 (first entry)  
 XX Streptococcus polynucleotide SEQ ID NO 10967.

XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;  
 KW group A streptococcus; Streptococcus pyogenes; antibacterial; gene;  
 KW antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.  
 OS Streptococcus sp.  
 XX WO200234771-A2.  
 XX 02-MAY-2002.  
 XX 29-OCT-2001; 2001WO-GB04789.  
 XX 27-OCT-2000; 2000GB-0026333.  
 PR 24-NOV-2000; 2000GB-0028727.  
 PR 07-MAR-2001; 2001GB-0005640.  
 XX (CHIR-) CHIRON SPA.  
 PA (GENO-) INST GENOMIC RES.  
 XX Telford J, Masighani V, Margarit Ros YI, Grandi G, Fraser C;  
 PI Tettelin H;  
 XX WPI; 2002-352536/38.  
 DR  
 XX New Streptococcus protein for the treatment or prevention of infection  
 PT or disease caused by Streptococcus bacteria, such as meningitis, and  
 PT for detecting a compound that binds to the protein -  
 XX  
 XX Claim 8; Page 4196-4488; 4525pp; English.  
 XX The invention relates to a protein (ABP25413-ABP30895) from group B  
 CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS  
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in  
 CC the specification. The proteins have antibacterial and antiinflammatory  
 CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and  
 CC antibodies that bind (I) are used in the manufacture of medicaments for  
 CC the treatment or prevention of infection or disease caused by  
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
 CC Nucleic acids encoding (I) are used to detect Streptococcus in a  
 CC biological sample. (I) is used to determine whether a compound binds to  
 CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be  
 CC used as a vaccine or diagnostic composition. The disease caused by  
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic  
 CC acid encoding (I) may be used to recombinantly produce (I) and may be  
 CC used in gene therapy. Antibodies to (I) are used for affinity  
 CC chromatography, immunoassays, and distinguishing/identifying  
 CC Streptococcus proteins.  
 XX  
 SQ Sequence 2155561 BP; 695741 A; 385320 C; 382958 G; 691542 T; 0 other;  
 Query Match 3.5%; Score 39.4; DB 24; Length 2155561;  
 Best Local Similarity 45.8%; Pred. No. 40;  
 Matches 136; Conservative 0; Mismatches 161; Indels 0; Gaps 0;  
 QY 40 GGTGACAAAGCTGTTGTTAAACAGATGCTCAGTTCCAGAAATTAAGGAGGTCACAGCC 99  
 DB 414490 GGCATAATGGTTATTAAGTATGATGATGATGATGATGATGATGATGATGATGAT 414431  
 QY 100 TTGGTGAAGTTGAGGCTGTTGCTGTTGATGATGATGATGATGATGATGATGATGATGAT 159  
 DB 414430 ATCAAAATTTGTTAGAGCCTGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 414371  
 QY 160 ATTGGTCCAGAGGTTCAATTTAGGATGATGATGATGATGATGATGATGATGATGATGAT 219  
 DB 414370 AAACGTGAAGTGGTTCTTTTTCAGGTCATGAGGCTATTGGTATCGTTAGGAAGTTGGT 414311  
 QY 220 CCAATGCTAGTACTGACTTGAAGTTGGAGATACCGGTTTCGGTTTGTTCACGGTGCT 279  
 DB 414310 ACTAAAGTAAGTACGACGTGTCATAAAGTGATTTTGTATTGTTCCCTTTACACATGCTGT 414251  
 QY 280 TCCCAACACAGATCCTTAAATGTTGATGTTGCTGAATATATGCCAGGTTTATCCACCT 336  
 |||||



```

RESULT 27
AA339814/c
ID AAX39814 standard; DNA; 847 BP.
XX
XX
AC AAX39814;
AC
XX
DT 02-JUL-1999 (first entry)
XX
XX Gastric cancer associated gene.
XX
XX Cancer associated antigen; diagnosis; research; treatment; human;
XX breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
XX prostate cancer; ss.
XX
OS Homo sapiens.
XX
XX WO9904265-A2.
XX
XX 28-JAN-1999.
XX
XX 15-JUL-1998; 98WO-US14679.
XX
XX 22-JUN-1998; 98US-0102322.
XX 17-JUL-1997; 97US-0896184.
XX 10-OCT-1997; 97US-0061599.
XX 10-OCT-1997; 97US-0061765.
XX 10-OCT-1997; 97US-0948705.
XX 11-OCT-1997; 97GB-0021697.
XX
XX (LUDM-) LUDWIG INST CANCER RES.
XX
XX Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;
XX Pfeundscher M, Sahin U, Scanlan MJ, Stockert E;
XX Tureci O;
XX
XX WPI; 1998-132448/11.
XX
XX New isolated cancer associated nucleic acids and polypeptides -
XX isolated using sera from cancer patients, used to develop products
XX for the diagnosis, monitoring or treatment of cancers
XX
XX Claim 67; Page 557; 787pp; English.
XX
XX The invention relates to a method for diagnosing a disorder characterised
XX by expression of a human cancer associated antigen precursor coded for by
XX a nucleic acid molecule (NAM). The method comprises: (a) contacting a
XX biological sample isolated from a subject with an agent that specifically
XX binds to the NAM, an expression product or a fragment of an expression
XX product complexed with an HLA molecule; and (b) determining the
XX interaction between the agent and the NAM or the expression product as a
XX determination of the disorder. The products and methods can be used in
XX the diagnosis, monitoring, research, or treatment of conditions
XX characterised by the expression of various cancer associated antigens.
XX The invention provides nucleic acid sequences and encoded polypeptides
XX which are cancer associated antigen precursors expressed in human breast
XX cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
XX lung cancer.
XX
XX Sequence 847 BP; 281 A; 159 C; 136 G; 263 T; 8 other;
XX
XX Query Match 3.5%; Score 38.4; DB 20; Length 847;
XX Best Local Similarity 46.6%; Pred. No. 3;
XX Matches 117; Conservative 0; Mismatches 134; Indels 0; Gaps 0;
XX
XX 837 AGAATCAGAAAGATATGTATAAATGTATATACCTTTGTTGTCATCGTCATCGTCA 896
XX
XX 773 AGAATTCGATTCGATTAATGNGAATAATTTATTTATATGATTTTGCATTCATCA 714
XX
XX 897 AGAATTCATTCGTCACACAGATTTCTGCTAGTCCAGATATCATGAGCCACAGT 956
XX
XX 713 TAACATTTAATTTGTTGAAGAAGTATTTTANTACCTATATAAATGAATCGAANT 654

```

---

```

QY 957 TAAATTCGTTAAGTTTATATAATCCACACTTAACACGGTGATATCCATCATATGAATAT 1016
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
653 TATTAAGTTCTTTAAACATAATTTACTCTTTTAAAAAATACCTGTTTCATATTTGTAAC 594
QY 1017 TAAAGTTTTCAGCAGCGCTTAGATGATGCTCCAGCTCTCACTGAGGTATATAAAGAAG 1076
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
593 CCATATTTTGGACACATTAATAAATGAATGTTGATATTACAGATATTTCATAACAAATG 534
QY 1077 TAAAAACAAAA 1087
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
533 GAAGATTATA 523
RESULT 28
AAF7507/c
ID AAF7507 standard; DNA; 486 BP.
XX
XX AAF7507;
AC
XX
XX 09-MAY-2001 (first entry)
XX
XX Polyglutamine tract coding sequence #2.
XX
XX Animal model; polyglutamine tract; neurodegenerative disorder; HDJ1;
XX heat shock protein 40; tetratricopeptide repeat protein 2; TPR2; stroke;
XX myeloid leukaemia factor 1; MLF; human; fruit fly; Alzheimer's disease;
XX Parkinson's disease; CJD; BSE; Huntington's disease; head trauma;
XX cancer; da.
XX
XX Drosophila sp.
OS
XX
XX WO200112238-A1.
XX
XX 22-FEB-2001.
XX
XX 14-AUG-2000; 2000WO-US22496.
XX
XX 12-AUG-1999; 99US-0148933.
XX 12-AUG-1999; 99US-0148934.
XX 18-JAN-2000; 2000US-0177047.
XX 19-MAY-2000; 2000US-0205720.
XX
XX (CALY ) CALIFORNIA INST OF TECHNOLOGY.
XX
XX Benzer S, Kazemi-Esfarjani P;
XX
XX WPI; 2001-147537/15.
XX
XX Identifying genes or other compounds that modulate polyglutamine
XX toxicity, useful for treating Alzheimer's disease, Parkinson's disease
XX and Creutzfeldt-Jakob disease -
XX
XX Disclosure; Fig 1A; 275pp; English.
XX
XX The present invention describes a method of screening for genes which
XX modulate polyglutamine toxicity using animal models with polyglutamine
XX sequences that cause toxicity in the animal. The model is preferably
XX Drosophila, and toxic polyglutamine sequences include the human and
XX Drosophila heat shock protein 40/HDJ1, tetratricopeptide repeat protein 2
XX (TPR2) and myeloid leukaemia factor 1 (MLF) genes. The model is useful
XX for identifying treatments for neurodegenerative and proliferative
XX disorders, including Alzheimer's disease, Parkinson's disease,
XX Creutzfeldt-Jacob's disease (CJD), bovine spongiform encephalopathy
XX (BSE), Huntington's disease, Machado-Joseph disease, spinocerebellar
XX ataxias, dentatorubralpallidoluysian atrophy, Kennedy's disease, stroke,
XX head trauma and cancer.
XX
XX Sequence 486 BP; 193 A; 164 C; 108 G; 21 T; 0 other;
XX
XX Query Match 3.4%; Score 38.2; DB 22; Length 486;
XX Best Local Similarity 43.7%; Pred. No. 2.7;
XX Matches 169; Conservative 0; Mismatches 218; Indels 0; Gaps 0;

```

QY 3 GTACGTTCCACCACTCAAAAGCGTCATCAITGAAAGGTGACAAAGCTGTGTGTTAAAC 62  
DB 446 GTCTGTTGCTGCTGTGTGTGCTGCTGTGTGTGCTGTGTGTGCTGTGTGTGCTG 387  
QY 63 AGATGCTCAGTCCAGATTAAGAGGAGGTACAGCCTTGGTGAAGGTTGAGGCTGTGC 122  
DB 386 TGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 327  
QY 123 TGGTAACCCACTGATGGAAGCATATGCTTAAAGATTGGTCCAGAAGTTCAATTCT 182  
DB 326 TGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 267  
QY 183 AGGATGTGACATTCCTGTACAGTGTCAAACTTGGACCAAAATGCTAGTACTGACTGAA 242  
DB 266 TGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 207  
QY 243 GGTGGAAGTACCGGTTTCGCTTTGTTTTCACGGTGTCCCAACAGATCCTAAAAATGG 302  
DB 206 TGTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 147  
QY 303 TGCATTTGCTGAATATGCCAGGTTTATCCACCTTTGTTTACAGAGTAACCTCACTCA 362  
DB 146 TGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 87  
QY 363 CTCAACTGCTGATGAAATTTCTGAAG 389  
DB 86 TGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 60

RESULT 29  
AAA68609  
ID AAA68609 standard; DNA; 43576 BP.  
AC AAA68609;  
XX  
XX 27-OCT-2000 (first entry)  
DT  
DE Bacteriophage 96 complete genome sequence.  
XX  
XX Bacteriophage; antimicrobial; genome; identification; antibacterial;  
KW Bacterial growth inhibition; bacterial infection; ds.  
XX  
XX Bacteriophage 96.  
OS  
XX WO200032825-A2.  
PN  
XX 08-JUN-2000.  
PD  
XX 03-DEC-1999; 99WO-IB02040.  
PF  
XX 03-DEC-1998; 98US-0110992.  
PR 03-JUN-1999; 99US-0326144.  
PR 28-SEP-1999; 99US-0407804.  
PR 30-SEP-1999; 99US-0157218.  
PR 01-DEC-1999; 99US-0168777.  
PR 02-DEC-1999; 99US-0454252.  
XX  
PA (PHAG-) PHAGETECH INC.  
XX  
XX Pelletier J, Gros P, Dubow M;  
PI  
XX WPI; 2000-412361/35.  
XX  
XX

PT Identifying a bacteriophage coding region for treating bacterial  
PT infections comprises identifying a nucleic acid encoding a product that  
PT inhibits bacteria when a bacteriophage infects a bacterium  
XX  
XX  
XX Disclosure; Page 190-198; 456pp; English.  
XX  
XX The present invention describes a method for identifying a bacteriophage  
CC coding region encoding a product active on an essential bacterial  
CC target. The method comprises identifying a nucleic acid sequence encoding  
CC a gene product that provides a bacteria-inhibiting function when an

CC uncharacterised bacteriophage infects a pathogenic bacterium. The  
CC compound active on a target of a bacteriophage inhibitor protein in a  
CC bacteria is used to treat or prevent a bacterial infection in an animal.  
CC AAA68243 to AAA69442 and AAB16523 to AAB16954 represent bacteriophage  
CC nucleotide and protein sequences which are used in the exemplification of  
CC the present invention.  
XX  
SQ Sequence 43576 BP; 16057 A; 6485 C; 8769 G; 12265 T; 0 other;  
Query Match 3.4%; Score 38; DB 21; Length 43576;  
Best Local Similarity 47.2%; Pred. No. 19;  
Matches 116; Conservative 0; Mismatches 130; Indels 0; Gaps 0;  
QY 764 ATAAAGTCACAGACAGATGCTACCTGCCACATTTATAGAGTGGTTCCAAATGACCATG 823  
DB 41253 ATGAGTTGCAATATATGTTGGTGGCTCAAGTCAATCAAGACATGTATCAAGATCTGCAT 41312  
QY 824 AAAGCAATTCCTGAAGAAATCAGAAAAAGATAATGTTAAAAATTGATATTACTTTTGTATC 883  
DB 41313 ACGTGTAAATGTAGGAATAATAAGATTATGATTATATGGTTAAATACACAGGGT 41372  
QY 884 GTGCATCTGTCAGAAATCTTATGGGTGCAACAAGATTTCTGCTAGTCCAGATATC 943  
DB 41373 ATGACATTTGTTCTAGAGATTTCATTTAGACGCGACGAGAGAAAATGCAAGTGGTGGCATG 41432  
QY 944 ATGAAGCCACAGTTAAATTCGTTAACTTTTAAATCCACACCTTAAACAAGGTGATATCC 1003  
DB 41433 TTATTATCAAGTCAATTCATTCGCGATCTATTGATAAAGTATACAGATGTTATTA 41492  
QY 1004 ATCAT 1009  
DB 41493 AAAATA 41498

RESULT 30  
ABL55643  
ID ABL55643 standard; DNA; 50000 BP.  
XX  
AC ABL55643;  
XX  
XX 01-JUL-2002 (first entry)  
DT  
XX AmEPV genome fragment#1.  
XX  
XX AmEPV; gene therapy; viral vector; chromosome mapping; gene mapping;  
KW genetic deficiency disorder; ds.  
XX  
XX Amsacta moorei entomopoxvirus.  
OS  
XX WO200212526-A2.  
PN  
XX 14-FEB-2002.  
PD  
XX 10-AUG-2001; 2001WO-US25287.  
PF  
XX 10-AUG-2000; 2000US-224479P.  
PR 14-SEP-2000; 2000US-0662254.  
PR  
XX (UYFL) UNIV FLORIDA.  
PA  
XX Moyer RW, Li Y, Bawden AL;  
PI  
XX WPI; 2002-227161/28.  
XX  
XX Novel recombinant entomopox virus vector useful for delivering  
PT polynucleotide encoding protein to vertebrate cell, comprises  
PT polynucleotide encoding protein operably linked with heterologous  
PT promoter sequence  
XX  
XX Disclosure; Page 125-150; 326pp; English.  
PS  
XX The invention relates to a recombinant entomopox virus (EPV) vector,  
CC comprising a polynucleotide encoding a protein operably linked with a

heterologous promoter sequence. The invention also concerns methods for providing gene therapy for genetic deficiency disorders. Vectors of the invention are useful for delivering a polynucleotide encoding a protein to a vertebrate cell, preferably a mammalian cell, such as a human cell. The vector is introduced into the vertebrate cell by infection in a viral particle, or by transfection, transduction, or injection either in vitro or in vivo. The vector is useful for the delivery and expression of biologically useful proteins in gene therapy protocols, and for delivering large DNA segments for engineering of vertebrate cells. Polynucleotides of the invention have applications in techniques such as their use as insertion sites for foreign genes of interest, hybridisation probes, for chromosome and gene mapping, in PCR technologies, and in the production of sense or antisense nucleic acids. Vectors of the invention provide for stable integration and expression of heterologous DNA in host cells, and are adapted for accepting large heterologous polynucleotide inserts which can be delivered in an infected or transformed cell and expressed in a stable fraction. The current sequence represents a fragment of the genome of the genus B entomopoxvirus from *Amblyomma americanum* (AmEPV).

Sequence 5000 BP; 19857 A; 4373 C; 4222 G; 21548 T; 0 other;

	Query Match	3.4%	Score 38	DB 24	Length 50000
CC	Best Local Similarity	47.5%	Fred. No. 21		
CC	Matches 113	Conservative 0	Mismatches 125	Indels 0	Gaps 0
Qy	859	AAAATTGATATTACTTTGTTGTCATCTGTCAGAAATTCATTGGTGCAACA	918		
Db	45093	AAAAGGTATAGTACCAATGATATTGTTACTTTTATTATAAAATAAATAATATCC	45152		
Qy	919	AGATTTCTGCTAGTCCAGAAATATCATGAAGCCACAGTTAAATTCGTTAAGTTTATAAT	978		
Db	45153	AAATAAAGCTCCCATGTTTAAAAATAATAAAGTCACAGTTAATATATAAATTAATATAT	45212		
Qy	979	CCACACCTTAACAACGGTGATATCCATCATATGATATTAAGTTTTCAGCAACGGCTTA	1038		
Db	45213	TTTTTCATTATAGTATTTTATAAAAAAATAATAATAAATGTTTAAACACAGATTTA	45272		
Qy	1039	GATGATGTCGCCAGCTCTCACCTGAAGGTATAAAGAGGTAAAAACAAAAATGTTAAGT	1096		
Db	45273	ACTAATGAGAGATATCAGAGCTGCTAATAAATTAATAAAAAATAATACTTGTGAAT	45330		

Search completed: January 10, 2004, 06:40:13  
Job time : 403 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 10, 2004, 06:26:03 ; Search time 88 Seconds  
(without alignments)  
5582.493 Million cell updates/sec

Title: US-10-081-644-1  
Perfect score: 1113  
Sequence: 1 atgcagttcccaaccactca.....agtatgttcaggattataa 1113

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 90 summaries

Database : Issued Patents NA.\*

1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	37.6	3.4	8654	4	US-08-961-527-98
c	36.6	3.3	1664976	4	US-08-916-421B-1
2	36.2	3.3	3255	4	US-09-601-198-108
3	36.2	3.3	580073	4	US-08-545-528D-1
4	36.2	3.3	1830121	4	US-09-557-884-1
5	36	3.2	1830121	4	US-09-557-884-1
6	36	3.2	1830121	4	US-09-557-884-1
7	35.8	3.2	840	3	US-09-042-771-1
8	35.6	3.2	849	4	US-09-107-532A-403
c	35.6	3.2	3312	4	US-09-601-198-58
9	35.6	3.2	3600	3	US-08-855-910-7
10	35.6	3.2	3792	2	US-08-992-334-1
11	35.6	3.2	3792	2	US-08-992-334-1
12	35.6	3.2	5234	2	US-08-992-334-2
13	35.6	3.2	5234	3	US-08-992-334-2
14	35.6	3.2	6722	2	US-08-992-334-3
15	35.6	3.2	6722	2	US-08-992-334-3
16	35.6	3.2	6722	3	US-08-302-752-3
17	35.2	3.2	1434	4	US-09-134-001C-650
18	35	3.1	927	4	US-09-328-352-2636
19	35	3.1	4860	4	US-09-328-352-2636
20	34.8	3.1	360	4	US-09-702-705-1408
21	34.8	3.1	360	4	US-09-702-705-1408
22	34.8	3.1	1044	4	US-09-736-457-1408
23	34.8	3.1	1044	4	US-09-134-001C-461
c	34.8	3.1	1053	4	US-09-107-532A-3324
24	34.8	3.1	1347	3	US-09/622
25	34.8	3.1	1347	4	US-08-165-922A-11
c	34.8	3.1	1374	4	US-09-601-198-131
26	34.8	3.1	4016	4	US-09-173-053-3
27	34.8	3.1	4016	4	US-09-173-053-3

#### ALIGNMENTS

RESULT 1  
US-08-961-527-98  
; Sequence 98, Application US/08961527  
; Patent No. 6420135  
; GENERAL INFORMATION:

Sequence 209, App  
Sequence 209, App  
Sequence 29, Appl  
Sequence 3644, Ap  
Sequence 1, Appl  
Sequence 1, Appl  
Sequence 14, Appl  
Sequence 1, Appl  
Sequence 2004, Ap  
Sequence 1, Appl  
Sequence 1, Appl  
Sequence 3, Appl  
Sequence 1, Appl  
Sequence 2, Appl  
Sequence 2, Appl  
Sequence 60, Appl  
Sequence 1, Appl  
Sequence 382, App  
Sequence 2274, Ap  
Sequence 21, Appl  
Sequence 12, Appl  
Sequence 1, Appl  
Sequence 1, Appl  
Sequence 218, App  
Sequence 1, Appl  
Sequence 3, Appl  
Sequence 157, App  
Sequence 157, App  
Sequence 157, App  
Sequence 278, App  
Sequence 278, App  
Sequence 278, App  
Sequence 3001, Ap  
Sequence 5, Appl  
Patent No. 5231168  
Sequence 50, Appl  
Sequence 212, App  
Sequence 1334, Ap  
Sequence 1062, Ap  
Sequence 7, Appl  
Sequence 55, Appl  
Sequence 3, Appl  
Sequence 1, Appl  
Sequence 1, Appl  
Sequence 48, Appl  
Sequence 48, Appl  
Sequence 48, Appl  
Sequence 48, Appl  
Sequence 30, Appl  
Sequence 4, Appl  
Sequence 124, App  
Sequence 224, App  
Sequence 9, Appl  
Sequence 1, Appl  
Sequence 69, Appl  
Sequence 77, Appl  
Sequence 95, Appl  
Sequence 5, Appl  
Sequence 5, Appl

APPLICANT: Charles Kunsch  
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences  
NUMBER OF SEQUENCES: 391  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/961,527  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Brookes, A. Anders  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PB340P1  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 98:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 8654 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-961-527-98

Query Match 3.4%; Score 37.6; DB 4; Length 8654;  
Best Local Similarity 54.3%; Pred. No. 1.8;  
Matches 76; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

Qy 788 CTGCCACATTATTAGAGTGGTCCCAATGACCATTGAAGCATTCCTCTGAAGAAATCAGAA 847  
Db 6452 CTTTCACATCTGTAGAAGTGATGCCAATTTGGATGATACCTATTGAAGTGGAAATCCGTG 6511

Qy 848 AGATATATGTAAATTCATATTACTTTGTGTATCGTCATCTGCTCAGAAATTCAT 907  
Db 6512 AGATGATATCAAGATGGATACCTTCGTTTCAGTGGTCCGGTGGACAAACGTCATA 6571

Qy 908 TGGGTGCAACAAGATTTCTCT 927  
Db 6572 AGGTTTCAACAGGTGTACGT 6591

RESULT 2  
US-08-916-421B-1/c  
Sequence 1, Application US/08916421B  
Patent No. 6503729  
GENERAL INFORMATION:  
APPLICANT: Bult et al.  
TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococcus jannaschii  
TITLE OF INVENTION: jannaschii  
FILE REFERENCE: PB275  
CURRENT APPLICATION NUMBER: US/08/916,421B  
CURRENT FILING DATE: 1997-08-22  
PRIOR APPLICATION NUMBER: US 60/024,428  
PRIOR FILING DATE: 1996-08-22  
NUMBER OF SEQ ID NOS: 3  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 1  
LENGTH: 1664976  
TYPE: DNA

ORGANISM: Methanococcus jannaschii  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (28222)..(28222)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (28257)..(28258)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (84773)..(84773)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (84808)..(84808)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (84812)..(84812)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (98120)..(98120)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (98159)..(98159)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (98239)..(98239)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (98266)..(98266)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (98343)..(98343)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (103998)..(103998)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (148948)..(148948)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (163385)..(163385)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (191989)..(191989)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (191995)..(191995)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (231980)..(231980)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (234187)..(234187)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (234220)..(234220)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (234814)..(234814)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (309398)..(309398)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (309418)..(309418)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (312837)..(312837)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (312993)..(312993)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (319226)..(319226)



/ OTHER INFORMATION: n equals a, t, c, or g  
/ NAME/KEY: misc feature  
/ LOCATION: (559167)..(559167)  
/ OTHER INFORMATION: n equals a, t, c, or g  
/ NAME/KEY: misc feature  
/ LOCATION: (559241)..(559241)  
/ OTHER INFORMATION: n equals a, t, c, or g  
/ NAME/KEY: misc feature  
/ LOCATION: (600992)..(600992)  
/ OTHER INFORMATION: n equals a, t, c, or g  
/ NAME/KEY: misc feature  
/ LOCATION: (622708)..(622708)  
/ OTHER INFORMATION: n equals a, t, c, or g  
/ NAME/KEY: misc feature  
/ LOCATION: (657081)..(657081)  
/ OTHER INFORMATION: n equals a, t, c, or g  
/ NAME/KEY: misc feature  
/ LOCATION: (657203)..(657203)  
/ OTHER INFORMATION: n equals a, t, c, or g  
/ NAME/KEY: misc feature  
/ LOCATION: (674435)..(674435)  
/ OTHER INFORMATION: n equals a, t, c, or g  
/ NAME/KEY: misc feature  
/ LOCATION: (682442)..(682442)  
/ OTHER INFORMATION: n equals a, t, c, or g  
/ NAME/KEY: misc feature  
/ LOCATION: (713652)..(713652)  
/ OTHER INFORMATION: n equals a, t, c, or g  
/ NAME/KEY: misc feature  
/ LOCATION: (741684)..(741684)  
/ OTHER INFORMATION: n equals a, t, c, or g  
/ NAME/KEY: misc feature  
/ LOCATION: (779455)..(779455)  
/ OTHER INFORMATION: n equals a, t, c, or g  
/ NAME/KEY: misc feature  
/ LOCATION: (779676)..(779676)  
/ OTHER INFORMATION: n equals a, t, c, or g  
/ NAME/KEY: misc feature  
/ LOCATION: (855539)..(855539)  
/ OTHER INFORMATION: n equals a, t, c, or g  
/ NAME/KEY: misc feature  
/ LOCATION: (871619)..(871619)  
/ OTHER INFORMATION: n equals a, t, c, or g  
/ NAME/KEY: misc feature  
/ LOCATION: (1084830)..(1084830)  
/ OTHER INFORMATION: n equals a, t, c, or g  
/ NAME/KEY: misc feature  
/ LOCATION: (1096846)..(1096846)  
/ OTHER INFORMATION: n equals a, t, c, or g  
/ NAME/KEY: misc feature  
/ LOCATION: (1119881)..(1119881)  
/ OTHER INFORMATION: n equals a, t, c, or g  
/ NAME/KEY: misc feature  
/ LOCATION: (1130881)..(1130881)  
/ OTHER INFORMATION: n equals a, t, c, or g  
/ NAME/KEY: misc feature  
/ LOCATION: (1310988)..(1310988)  
/ OTHER INFORMATION: n equals a, t, c, or g  
/ NAME/KEY: misc feature  
/ LOCATION: (1313224)..(1313224)  
/ OTHER INFORMATION: n equals a, t, c, or g  
/ NAME/KEY: misc feature  
/ LOCATION: (1349473)..(1349473)  
/ OTHER INFORMATION: n equals a, t, c, or g  
/ NAME/KEY: misc feature  
/ LOCATION: (1349491)..(1349491)  
/ OTHER INFORMATION: n equals a, t, c, or g  
/ NAME/KEY: misc feature  
/ LOCATION: (1470091)..(1470091)  
/ OTHER INFORMATION: n equals a, t, c, or g  
/ NAME/KEY: misc feature  
/ LOCATION: (1569020)..(1569020)  
/ OTHER INFORMATION: n equals a, t, c, or g

/ NAME/KEY: misc feature  
/ LOCATION: (1602912)..(1602912)  
/ OTHER INFORMATION: n equals a, t, c, or g  
/ NAME/KEY: misc feature  
/ LOCATION: (1603734)..(1603734)  
/ OTHER INFORMATION: n equals a, t, c, or g  
/ NAME/KEY: misc feature  
/ LOCATION: (1637998)..(1637998)  
/ OTHER INFORMATION: n equals a, t, c, or g  
/ NAME/KEY: misc feature  
/ LOCATION: (1664855)..(1664855)  
/ OTHER INFORMATION: n equals a, t, c, or g  
US-08-916-421B-1  
  
Query Match 3.3%; Score 36.6; DB 4; Length 1664976;  
Best Local Similarity 49.7%; Pred. No. 27;  
Matches 93; Conservative 0; Mismatches 94; Indels 0; Gaps 0;  
  
QY 723 TATTGACGCTGTGGGAAGCGAAGATAGTAGTATCCCCGAGGCGCTATAAAGTCACAGCAGATAG 782  
Db 1311832 TTTTGAATCTGATAAAGTAGGAGAGATTTTAAATAAATAGATAAAGATACTCTAGTTAT 1311773  
  
QY 783 TCTACCTGCCACATTATTAGAAAGTGGTCCCAATGACCATTGAAGCATTCCTGAAGAAAT 842  
Db 1311772 TATCACTGGTTCACCAATCGATTTGAAAGTTATTGAACCTTTAAAAAAGTTAT 1311713  
  
QY 843 CAGAAAGATAATGTTAAAAATTGATATTACTTTGTTGATCGTGCAATCTGGTCAAGAAAT 902  
Db 1311712 TAATCAGCTAAATAAATAAAGAGCTATTGCTTTGTTGATGATGCTCTGGAGCGAGAGT 1311653  
  
QY 903 TCTATTG 909  
Db 1311652 TAGGTTG 1311646  
  
RESULT 3  
US-09-601-198-108/c  
; Sequence 108, Application US/09601198  
; Patent No. 6531583  
; GENERAL INFORMATION:  
; APPLICANT: Caswell, Gail H.  
; APPLICANT: Chen, Ellison Y.  
; APPLICANT: Glass, Jennifer S.  
; APPLICANT: Glass, John I.  
; APPLICANT: Heiner, Cheryl R.  
; APPLICANT: Leftkowitz, Elliot  
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND METHOD FOR DETECTING UREAPLASMA  
; TITLE OF INVENTION: UREALYTICUM  
; FILE REFERENCE: UAB-13452/22  
; CURRENT APPLICATION NUMBER: US/09/601,198  
; CURRENT FILING DATE: 2000-12-08  
; PRIOR APPLICATION NUMBER: 60/073,189  
; PRIOR FILING DATE: 1998-01-30  
; NUMBER OF SEQ ID NOS: 181  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 108  
; LENGTH: 3255  
; TYPE: DNA  
; ORGANISM: Ureaplasma urealyticum  
US-09-601-198-108

Query Match 3.3%; Score 36.2; DB 4; Length 3255;  
Best Local Similarity 45.8%; Pred. No. 2, 9;  
Matches 125; Conservative 0; Mismatches 148; Indels 0; Gaps 0;  
  
QY 829 ATTCCTGAAGAAATCAGAAAAGATAATGTTAAATTTGATATTACTTTGTTGTTGTCGTGCA 888  
Db 3225 ATAGGTGTTGTATCAACAATAATTTCTTGTTCATAAATATATATCATTTCTAATTGTTCT 3166  
  
QY 889 TCTGGTCAAGAAATTTCTATTGGGTGCAACAAGATTTTCCTGCTAGTCCAGATATCATGAA 948  
Db 3165 TATAACAAAAAACATTTTCGTCTTCTACAATTTACGAATCAAGATATAAATAATTTAGAC 3106





```

; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (B) LOCATION 1...849
; SEQUENCE DESCRIPTION: SEQ ID NO: 403:
US-09-107-532A-403

Query Match          3.2%; Score 35.6; DB 4; Length 849;
Best Local Similarity 46.7%; Pred. No. 2.4;
Matches 113; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

Qy 867 TATTACTTTTGTGATCGTGCATCTGGTCAAGAAATCTATTGGTGCACAAAGATTTC 926
Db 543 TGTTCCAGTCGTTGATGTTTCACTTAAGTGGTTCAGTATTGAAACAAAGTAAC 602
Qy 927 TGTGATGTCAGATPATCATGAAGCCACAGTGTAAATTCGTTAAAGTTTATAATCCACCT 986
Db 603 TGCTGATCAAGTAACGAAGCAATCAAAAAACACACAGTTGACAACTCTTCATTGGGTA 662
Qy 987 TAACAAGGTCATATCCATCATATGAATATTAAGTTTTCAGCAACGGCTAGATGATGT 1046
Db 663 TGATGACCGTCAAAATCGTTTTCAGGATGTCATGGAACCACTCAAGGTTTCGATCTTGA 722
Qy 1047 CCCAGCTCTCACTGAAGGTATATAAAGAGGTAAACAAACAAATGTTAAGTATGTGCCAG 1106
Db 723 CCCAAGCAACAGAGTAACAACTGCTGGTGATTTCCAACTGTTAAACTGTTGCATG 782
Qy 1107 GT 1108
Db 783 GT 784

RESULT 9
US-09-601-198-58/c
; Sequence 58, Application US/09601198
; Patent No. 6531583
; GENERAL INFORMATION:
; APPLICANT: Cassell, Gail H.
; APPLICANT: Chen, Ellison Y.
; APPLICANT: Glass, Jennifer S.
; APPLICANT: Glass, John I.
; APPLICANT: Heiner, Cheryl R.
; APPLICANT: Lefkowitz, Elliot
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND METHOD FOR DETECTING UREAPLASMA
; FILE REFERENCE: UAB-13452/22
; CURRENT APPLICATION NUMBER: US/09/601.198
; CURRENT FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/073,189
; PRIOR FILING DATE: 1998-01-30
; NUMBER OF SEQ ID NOS: 181
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 58
; LENGTH: 3312
; TYPE: DNA
; ORGANISM: Ureaplasma urealyticum
US-09-601-198-58

```

```

Query Match          3.2%; Score 35.6; DB 4; Length 3312;
Best Local Similarity 48.1%; Pred. No. 4.2;
Matches 101; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

Qy 894 TCAAGAAATCTATTGGTGCACAAAGATTTCCTGCTAGTCCAGATATCATGAAGCCAC 953
Db 2893 TAATGAAGATTAAATTTACAACCATCACTTCCTAATACACAAAGATTGATCATTTAGA 2834
Qy 954 AGTTAAATTCGTTAAAGTTTATAATCCACACCTTAACACGGTCATATCCATCATATGAA 1013
Db 2833 AATTAAATATTCGTTTTCATAATATAAACAACCAAGTTGATATAAAGATTAAAGTT 2774
Qy 1014 TATTAAGTTTTCAGCAACCGCTTAGATGATGTGCCAGCTCTCACTGAAGGTATAAAGA 1073

```

```

Db 2773 AAAAAAGAGTTTCTTTTAAAGACGATGCTATTCATATTTTAGTTTAAAAAATATCAAAA 2714
Qy 1074 AGTAAAAACAAAAATGTTAAGTATGTTGC 1103
Db 2713 AACTAATACCAAAATCACTACTAGTGATGC 2684

RESULT 10
US-08-855-910-7
; Sequence 7, Application US/08855910
; Patent No. 6221640
; GENERAL INFORMATION:
; APPLICANT: Tao, Jianehi
; APPLICANT: Sasanfar, Mandana
; APPLICANT: Gallant, Paul L.
; APPLICANT: Shen, Xiaoyu
; APPLICANT: Avruch, Anthony S.
; APPLICANT: Yu, Russell V.
; APPLICANT: Nair, Shamila
; TITLE OF INVENTION: ENTEROCOCCAL AMINOACYL-tRNA SYNTHETASE
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACIDS AND STRAINS COMPRISING SAME
; NUMBER OF SEQUENCES: 71
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/855,910
; FILING DATE: 14-MAY-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Brook, David E.
; REGISTRATION NUMBER: 22,592
; REFERENCE/DOCKET NUMBER: CPI95-08
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781) 861-6240
; TELEFAX: (781) 861-9540
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3600 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 85..1128
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1139..3559
US-08-855-910-7

Query Match          3.2%; Score 35.6; DB 3; Length 3600;
Best Local Similarity 51.9%; Pred. No. 4.4;
Matches 80; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

Qy 708 AAACCTGCAACATGTTATTGACGCTGTGGAGACGAGATAGTATCCCCGAGCCCTATAA 767
Db 3223 AAACTGCAACATGTTATTGACGCTGTGGAGACGAGATAGTATGAAAGCAGTTTCTAA 3282
Qy 768 AGTCACAGCAGATGCTCTACCTGCCACATTTATGAGTGGTTCCTCAATGACCATTCGAAG 827
Db 3283 ATTCCAGCAGATGCTCTCGAGATATTGCTTTATTAGTCGATGAACAGTTTACCANTCAAG 3342
Qy 828 CATTCCTGAGAAATCAGAAAAAGATAAATGTTAAA 861

```

Db 3343 ACTAGTAAACTATTTTCAGATAACCGAGGTAAA 3376

RESULT 11  
US-08-992-334-1  
; Sequence 1, Application US/08992334  
; Patent No. 5919678  
; GENERAL INFORMATION:  
; APPLICANT: Gruss, Alexandra  
; APPLICANT: Maguin, Emmanuelle  
; TITLE OF INVENTION: METHODS FOR USING A TEMPERATURE-SENSITIVE  
; TITLE OF INVENTION: PLASMID  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Christie Parker & Hale, LLP  
; STREET: 350 West Colorado Boulevard, Suite 500  
; CITY: Pasadena  
; STATE: California  
; COUNTRY: United States  
; ZIP: 91105  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/992,334  
; FILING DATE: 17-DEC-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/302,752  
; FILING DATE: 24-DEC-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/FR93/00248  
; FILING DATE: 12-MAR-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: FR FR92/03034  
; FILING DATE: 13-MAR-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Prout, D. Bruce  
; REGISTRATION NUMBER: 20958  
; REFERENCE/DOCKET NUMBER: C93:31779  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (626) 795-9900  
; TELEFAX: (626) 577-8800  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3792 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: both  
; TOPOLOGY: circular  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: YES  
; IMMEDIATE SOURCE:  
; CLONE: pg+host4  
US-08-992-334-1

Query Match 3.2%; Score 35.6; DB 2; Length 3792;  
Best Local Similarity 47.3%; Pred. No. 4.5;  
Matches 107; Conservative 0; Mismatches 119; Indels 0; Gaps 0;

Qy 820 ATTGAAGCATTCCTGGAAGAAATCTAATGCGTGCACAAAGATTCTCTGCTAGTCCAGAA 879  
Db 2229 ATTGACGAATCTGTAACTAAGAGCGTTAGGAACAGATTAAAGCGAAATTTGGG 2288

Qy 880 TATCGTGCATCGTGCAGAAATCTAATGCGTGCACAAAGATTCTCTGCTAGTCCAGAA 939  
Db 2289 AATAGTTCAGTTGCTCATGTTGAGTACTTGTATATATCAAGGTTTCATATGAATTTG 2348

Qy 940 TATCATGAAGCCACAGTTAAATTCGTTTAAATTCACACCTTAAACACGGTGAT 999  
Db 2349 ACTCATGAATCAAGGACGCTATTGCTAAGATAAATATATACGACAAAGATATT 2408

Qy 1000 ATCCATCATATGAATATTAAAGTTTTCAGCAACGGCTTAGATGATG 1045  
Db 2409 TTGAACATTAAATGATTTTGATATTGACCGCTATATAACACTTGATG 2454

RESULT 12  
US-08-302-752-1  
; Sequence 1, Application US/08302752  
; Patent No. 6025190  
; GENERAL INFORMATION:  
; APPLICANT:  
; TITLE OF INVENTION: THERMOSENSIBLE PLASMID  
; NUMBER OF SEQUENCES: 3  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (BPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/302,752  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: FR 9203034  
; FILING DATE: 13-MAR-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: WO FR/93/00248  
; FILING DATE: 12-MAR-1993  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3792 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-302-752-1

Query Match 3.2%; Score 35.6; DB 3; Length 3792;  
Best Local Similarity 47.3%; Pred. No. 4.5;  
Matches 107; Conservative 0; Mismatches 119; Indels 0; Gaps 0;

Qy 820 ATTGAAGCATTCCTGGAAGAAATCTAATGCGTGCACAAAGATAATGTTAAATTTGATTTACTTTGTTG 879  
Db 2229 ATTGACGAATCTGTAACTAAGAGCGTTAGGAACAGATTAAAGCGAAATTTGGG 2288

Qy 880 TATCGTGCATCGTGCAGAAATCTAATGCGTGCACAAAGATTCTCTGCTAGTCCAGAA 939  
Db 2289 AATAGTTCAGTTGCTCATGTTGAGTACTTGTATATATCAAGGTTTCATATGAATTTG 2348

Qy 940 TATCATGAAGCCACAGTTAAATTCGTTTAAATTCACACCTTAAACACGGTGAT 999  
Db 2349 ACTCATGAATCAAGGACGCTATTGCTAAGATAAATATATACGACAAAGATATT 2408

Qy 1000 ATCCATCATATGAATATTAAAGTTTTCAGCAACGGCTTAGATGATG 1045  
Db 2409 TTGAACATTAAATGATTTTGATATTGACCGCTATATAACACTTGATG 2454

RESULT 13  
US-08-992-334-2  
; Sequence 2, Application US/08992334  
; Patent No. 5919678  
; GENERAL INFORMATION:  
; APPLICANT: Gruss, Alexandra  
; APPLICANT: Maguin, Emmanuelle  
; TITLE OF INVENTION: METHODS FOR USING A TEMPERATURE-SENSITIVE  
; TITLE OF INVENTION: PLASMID  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Christie Parker & Hale, LLP  
; STREET: 350 West Colorado Boulevard, Suite 500  
; CITY: Pasadena

```
; STATE: California
; COUNTRY: United States
; ZIP: 91105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/992,334
; FILING DATE: 17-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,752
; FILING DATE: 24-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/FR93/00248
; FILING DATE: 12-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 92/03034
; FILING DATE: 13-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Prout, D. Bruce
; REGISTRATION NUMBER: 20958
; REFERENCE/DOCKET NUMBER: C93:31779
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (626) 795-9900
; TELEFAX: (626) 577-8800
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5234 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; US-08-992-334-2

Query Match 3.2%; Score 35.6; DB 2; Length 5234;
Best Local Similarity 47.3%; Pred. No. 5.1;
Matches 107; Conservative 0; Mismatches 119; Indels 0; Gaps 0;

QY 820 ATTGAAGCATTCTCTGAAGAAATCAGAAAGATAATGTTAAATTTGATATTTGTTG 879
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3671 ATTGCAGAAATCTGTAAACAATAGAAAGCGTTAGGAACAAGATTAAAGCGAAATTTGGG 3730
QY 880 TATCGTGCATCTGTCACAGAAATTCATTGGGTGCAACAAGATTTCTCTGCTAGTCCAGAA 939
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3731 AATAGTTTCAGTTGCTCATGTTGAGATACCTTGATTTATATCAAGGTTTCATATGAATTTG 3790
QY 940 TATCATGAAGCCACAGTTAAATTCGTTAAAGTTTATAAATCCACACCTTAACAACGGTGAT 999
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3791 ACTCATGAATCAAGGACGCTATTGCTAAGAATAAACAATATATACGACAAAAAGATATT 3850
QY 1000 ATCCATCATATGAATTAAGTTTTCAGCAACGGCTTAGATGATG 1045
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3851 TTGAACATTAAATGATTTTGATATTGACCGCTATATAACACTTGATG 3896

RESULT 15
US-08-992-334-3
; Sequence 3, Application US/08992334
; Patent No. 5919678
; GENERAL INFORMATION:
; APPLICANT: Gruss, Alexandra
; APPLICANT: Mequin, Emmanuelle
; TITLE OF INVENTION: METHODS FOR USING A TEMPERATURE-SENSITIVE
; TITLE OF INVENTION: PLASMID
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christle Parker & Hale, LLP
; STREET: 350 West Colorado Boulevard, Suite 500
; CITY: Pasadena
; STATE: California
; COUNTRY: United States
; ZIP: 91105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/992,334
; FILING DATE: 17-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,752
; FILING DATE: 24-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/FR93/00248
; FILING DATE: 12-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 92/03034
; FILING DATE: 13-MAR-1992
```

ATTORNEY/AGENT INFORMATION:  
NAME: Prout, D. Bruce  
REGISTRATION NUMBER: 20958  
REFERENCE/DOCKET NUMBER: C93:31779  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (626) 795-9900  
TELEFAX: (626) 577-8800  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6722 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: circular  
MOLECULE TYPE: DNA (genomic)  
US-08-992-334-3

Query Match 3.2%; Score 35.6; DB 2; Length 6722;  
Best Local Similarity 47.3%; Pred. No. 5.6;  
Matches 107; Conservative 0; Mismatches 119; Indels 0; Gaps 0;

QY 820 ATTGAAGCAATTCCTGAAGAAATCAGAAAAGATAAATGTTAAATTCATATTACTTTGTTG 879  
DB 5159 ATTGCACGAATCCTGTAAACAATAGAAAGCGTTAGGAACAAGATTAAAGCAAAATTTGGGG 5218  
QY 880 TATCGTGCACTGGTCAAGAAATTCATTGGGGTGCACAGATTTTCCTCTAGTCCAGAA 939  
DB 5219 AATAGTTCAGTTCCTCATCTTGAGATACCTTGATTTATATCAAGGTTTCATATGAATATTG 5278  
QY 940 TATCATGAAGCCACAGTTAAATTCGTTAAAGTTTATAAATCCACACCTTAACACCGGTGAT 999  
DB 5279 ACTCATGAATCAAGGACGCTATTGCTAAGAATAAACAATATATACGACAAAAGATATT 5338  
QY 1000 ATCCATCATATGAATTAAGTTTTCAGCAACGGCTTAGATGATG 1045  
DB 5339 TTGAACATTAAATGATTTTGATATTGACCGCTATATAACACCTTGATG 5384

Query Match 3.2%; Score 35.6; DB 3; Length 6722;  
Best Local Similarity 47.3%; Pred. No. 5.6;  
Matches 107; Conservative 0; Mismatches 119; Indels 0; Gaps 0;

QY 820 ATTGAAGCAATTCCTGAAGAAATCAGAAAAGATAAATGTTAAATTCATATTACTTTGTTG 879  
DB 5159 ATTGCACGAATCCTGTAAACAATAGAAAGCGTTAGGAACAAGATTAAAGCAAAATTTGGGG 5218  
QY 880 TATCGTGCACTGGTCAAGAAATTCATTGGGGTGCACAGATTTTCCTCTAGTCCAGAA 939  
DB 5219 AATAGTTCAGTTCCTCATCTTGAGATACCTTGATTTATATCAAGGTTTCATATGAATATTG 5278  
QY 940 TATCATGAAGCCACAGTTAAATTCGTTAAAGTTTATAAATCCACACCTTAACACCGGTGAT 999  
DB 5279 ACTCATGAATCAAGGACGCTATTGCTAAGAATAAACAATATATACGACAAAAGATATT 5338  
QY 1000 ATCCATCATATGAATTAAGTTTTCAGCAACGGCTTAGATGATG 1045  
DB 5339 TTGAACATTAAATGATTTTGATATTGACCGCTATATAACACCTTGATG 5384

RESULT 16  
US-08-302-752-3  
Sequence 3, Application US/08302752  
Patent No. 6025190  
GENERAL INFORMATION:  
APPLICANT:  
TITLE OF INVENTION: THERMOSENSIBLE PLASMID  
NUMBER OF SEQUENCES: 3  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25 (BPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/302,752  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 9203034  
FILING DATE: 13-MAR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO FR/93/00248  
FILING DATE: 12-MAR-1993  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6722 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-302-752-3

Query Match 3.2%; Score 35.6; DB 3; Length 6722;  
Best Local Similarity 47.3%; Pred. No. 5.6;  
Matches 107; Conservative 0; Mismatches 119; Indels 0; Gaps 0;

QY 820 ATTGAAGCAATTCCTGAAGAAATCAGAAAAGATAAATGTTAAATTCATATTACTTTGTTG 879  
DB 5159 ATTGCACGAATCCTGTAAACAATAGAAAGCGTTAGGAACAAGATTAAAGCAAAATTTGGGG 5218  
QY 880 TATCGTGCACTGGTCAAGAAATTCATTGGGGTGCACAGATTTTCCTCTAGTCCAGAA 939  
DB 5219 AATAGTTCAGTTCCTCATCTTGAGATACCTTGATTTATATCAAGGTTTCATATGAATATTG 5278  
QY 940 TATCATGAAGCCACAGTTAAATTCGTTAAAGTTTATAAATCCACACCTTAACACCGGTGAT 999  
DB 5279 ACTCATGAATCAAGGACGCTATTGCTAAGAATAAACAATATATACGACAAAAGATATT 5338  
QY 1000 ATCCATCATATGAATTAAGTTTTCAGCAACGGCTTAGATGATG 1045  
DB 5339 TTGAACATTAAATGATTTTGATATTGACCGCTATATAACACCTTGATG 5384

RESULT 17  
US-09-134-001C-650  
Sequence 650, Application US/09134001C  
Patent No. 6380370  
GENERAL INFORMATION:  
APPLICANT: Lynn Doucette-Stamm et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
FILE REFERENCE: GTC-007  
CURRENT APPLICATION NUMBER: US/09/134,001C  
CURRENT FILING DATE: 1998-08-13  
PRIOR APPLICATION NUMBER: US 60/064,964  
PRIOR FILING DATE: 1997-11-08  
PRIOR APPLICATION NUMBER: US 60/055,779  
PRIOR FILING DATE: 1997-08-14  
NUMBER OF SEQ ID NOS: 5674  
SEQ ID NO 650  
LENGTH: 1434  
TYPE: DNA  
ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-650

Query Match 3.2%; Score 35.2; DB 4; Length 1434;  
Best Local Similarity 47.7%; Pred. No. 3.9;  
Matches 103; Conservative 0; Mismatches 113; Indels 0; Gaps 0;

QY 774 AGCAGATAGTCTACCTGCCACACATATTAGAAGTGGTTCCAAATGACCATTTGAAAGCATTC 833  
DB 552 AGCTTTTATTAGGTTCTCTCTCTAGAAAATAATACAAAACAAAGACATTTGTTCT 611  
QY 834 TGAAGAAATCAGAAAAGATAAATGTTAAATTTGATATATTCTTTGTTGATCGTCATCTGG 893  
DB 612 TGATAAAATTTCTGTAGTTATTCTACTTTTCGGCTTTGGTTAAATTTTATTCGCAITTAG 671  
QY 894 TCNAGAAATTCATTGGGTGCACACAGATTTCTGCTAGTCCAGATATATCATGAAGCCAC 953  
DB 672 TAGCGTGAGTAGCTTTGGGTATCACCTCTCTCATTTGTGACATTTGTACTCGGTAT 731  
QY 954 AGTTAAATTCGTTAAGTTTATAAATCCACACCTTTAA 989  
DB 732 AGCCATTATCAATTAATTTTACACGCGTCAGCTTAA 767

RESULT 18  
US-09-328-352-2636  
Sequence 2636, Application US/09328352  
Patent No. 6562958  
GENERAL INFORMATION:  
APPLICANT: Gary L. Breton et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
FILE REFERENCE: GTC99-03PA  
CURRENT APPLICATION NUMBER: US/09/328,352  
CURRENT FILING DATE: 1999-06-04  
NUMBER OF SEQ ID NOS: 8252  
SEQ ID NO 2636

```
; LENGTH: 927
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-2636

Query Match      3.1%; Score 35; DB 4; Length 927;
Best Local Similarity 45.2%; Pred. No. 3.7;
Matches 128; Conservative 0; Mismatches 155; Indels 0; Gaps 0;

QY  31 ATCATGGAAGTGCAAGCTGTTGTTAAACAGATCTCTCAGTTCAGAAATTAAGGAG 90
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  529 ATGAATGAAAAGGCCAAGCAATTAATGTACAAATATTTATGTTAAAGACCATGCTGAC 588
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  91 GGTACAGCCTTGGTGAAGTTGAGGCTGTTGCTGTAACCAACTGATTGGAAGCATATT 150
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  589 TCITTTGCCAATGATGGCATCTGAGCGGCTGCTGCATTTGTAATGGATGATATATTTTA 648
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  151 GCTTATAAGATTTGGTCCAGAAGTTCAATTTCTAGGATGTGACATTTGCTGTACAGTTGTC 210
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  649 GCAGGCTTAATTGCTAAGTCTTCTACACCTAAAGCATTTGCAATTTGTTGCTCCAGTACTT 708
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  211 AACTTGGACCAATAGCTAGTACTGACTTGAAGTTGGAGATACCGGTTTCGGTTTGT 270
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  709 TCATCAGAACCCTTATGCTATGATTTGCAAAAGATGATCCAAATTTAAAGCAATTGCT 768
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  271 CAGGTCCTTCCCAACAGATCCTTAAATAATGTTGCTGCTG 313
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  769 GACGTAAGTAATAACTTATGGAATACTGGTCAAGTGGATG 811
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 19
US-09-328-352-3221
; Sequence 3221, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 3221
; LENGTH: 4860
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-3221

Query Match      3.1%; Score 35; DB 4; Length 4860;
Best Local Similarity 57.9%; Pred. No. 7.2;
Matches 62; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY  150 TGCTTATAAGATTTGGTCCAGAAGTTCAATTTCTAGGATGTGACATTTGCTGTACAGTTGT 209
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  855 TGACGATAAAACTGCGCAATAGTTTAAACAGTTGCAATTTGATACAGCGGTAAAGTAGT 914
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  210 CAACTTGGACCAATCTAGTACTGACTTGAAGTTGGAGATACCG 256
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  915 TAAAGGTGTAAGTGAAGCCCAATGCTGTTGTGACAGTTTAAAGATGCTG 961
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 20
US-09-702-705-1408
; Sequence 1408, Application US/09702705
; Patent No. 6504010
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darrick
; APPLICANT: Retter, Marc
```

```
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Liqun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C14
; CURRENT APPLICATION NUMBER: US/09/702,705
; CURRENT FILING DATE: 2000-10-30
; NUMBER OF SEQ ID NOS: 1833
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1408
; LENGTH: 360
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-702-705-1408

Query Match      3.1%; Score 34.8; DB 4; Length 360;
Best Local Similarity 51.3%; Pred. No. 2.8;
Matches 81; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY  927 TGCTAGTCCAGAATATCATGAAGCCACAGTTAAATTCGTTAAAGTTTATATAATCCACACCT 986
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  200 TGGAGTCCAAATTCATGCTGCTGTGAGTGTAAATAATATATGTTTCTTAAGACCGTGTGT 259
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  987 TAACACGGTGTATCCATCATATGATTAATAAGTTTTCAGCAACGGCTTAGATGATGT 1046
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  260 GAAAGATATAAAATATCTCATTAAAAATGTTTATATTGAGTACATGTTGAAATAATTT 319
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  1047 CCAGCTCTCACTGAAGGTATAAAGAGGTATAAACA 1084
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  320 TATATTGTGACACATTTGTTAAAAATATTTAAA 357
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 21
US-09-736-457-1408
; Sequence 1408, Application US/09736457
; Patent No. 6509448
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darrick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Liqun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C15
; CURRENT APPLICATION NUMBER: US/09/736,457
; CURRENT FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 1864
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1408
; LENGTH: 360
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-736-457-1408

Query Match      3.1%; Score 34.8; DB 4; Length 360;
Best Local Similarity 51.3%; Pred. No. 2.8;
Matches 81; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY  927 TGCTAGTCCAGAATATCATGAAGCCACAGTTAAATTCGTTAAAGTTTATATAATCCACACCT 986
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  200 TGGAGTCCAAATTCATGCTGCTGTGAGTGTAAATAATATATGTTTCTTAAGACCGTGTGT 259
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  987 TAACACGGTGTATCCATCATATGATTAATAAGTTTTCAGCAACGGCTTAGATGATGT 1046
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  260 GAAAGATATAAAATATCTCATTAAAAATGTTTATATTGAGTACATGTTGAAATAATTT 319
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  1047 CCAGCTCTCACTGAAGGTATAAAGAGGTATAAACA 1084
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```



Db 320 TATATTGTGACACATTGTGTTAAATAAATAATATAAAA 357  
RESULT 22  
US-09-134-001C-461  
; Sequence 461, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134.001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO. 461  
; LENGTH: 1044  
; TYPE: DNA  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-461  
Query Match 3.1%; Score 34.8; DB 4; Length 1044;  
Best Local Similarity 45.6%; Pred. No. 4.4;  
Matches 123; Conservative 0; Mismatches 147; Indels 0; Gaps 0;  
QY 499 CCGCAACATCTCATTCATTATTTGATTTGGGTGGTCTACAGCAGTGGGTCAACACTA 558  
Db 463 CCTAAGAAGGTGAACAGTAGTATTATTCAGCAGCTTCAGGTGCAGTTGTTCAAGTTGTG 522  
QY 559 ATCCAGTTCGCCAACATCAATGCTTATCTACTAAGATTGTAAGTCTTCTTCTTAAAG 618  
Db 523 GGGCAATTCGGAAGCTTAAAGGTTGCAGAGTCGTTGGTATAGCTGTTGGAGATAAAAA 582  
QY 619 CATGMAAAGCTTTTAAAGTCTTATGGTGTGATGATCTTTGACTATCATGATGCAGGC 678  
Db 583 GTCAACTATCTAAAAATGAATCTCGTTTGTGCTGCTGATCGATTCGATCAAAAAAGATAAT 642  
QY 679 GTTATTGAGCAGATCAATCGAAGTATCCAAACCTGCAACATGTTATTTGACGCTGTGGGA 738  
Db 643 TTCCTGAAGCGTTAAAGAAGCGGTGCCTAACGGTATAGATGCTACTTTCGAAAAATGTA 702  
QY 739 AGCGAGATAGTATCCCGGAGGCTATAAA 768  
Db 703 GGTGGATATATTGGCGATGAAGTCTTCAAA 732  
RESULT 23  
US-09-107-532A-3324/C  
; Sequence 3324, Application US/09107532A  
; Patent No. 6583275  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; TITLE OF INVENTION: ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
; NUMBER OF SEQUENCES: 7310  
; CORRESPONDENCE ADDRESS:  
; STREET: 100 Beaver Street  
; CITY: Waltham  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02354  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: CD-ROM ISO9660  
; COMPUTER: PC  
; OPERATING SYSTEM: <Unknown>  
; SOFTWARE: ASCII  
; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/107,532A  
; FILING DATE: 30-Jun-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/085,598  
; FILING DATE: 14 May 1998  
; APPLICATION NUMBER: 60/051571  
; FILING DATE: July 2, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ariniello, Pamela Deneke  
; REGISTRATION NUMBER: 40,489  
; REFERENCE/DOCKET NUMBER: GTC-012  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (781)893-5007  
; TELEFAX: (781)893-8277  
; INFORMATION FOR SEQ ID NO: 3324:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1053 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: circular  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Enterococcus faecium  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (8) LOCATION 1...1053  
; SEQUENCE DESCRIPTION: SEQ ID NO: 3324:  
US-09-107-532A-3324  
Query Match 3.1%; Score 34.8; DB 4; Length 1053;  
Best Local Similarity 52.0%; Pred. No. 4.4;  
Matches 78; Conservative 0; Mismatches 72; Indels 0; Gaps 0;  
QY 920 GATTTCTGCTAGTCCAGAATATCATGAAGCCACAGTTAAATTCGTTAAGTTTATAAATC 979  
Db 980 GTTCCACAGTCATCAAGACAGTCATGAACCAAGGAACCTTCTTTCACTTCTTTGATC 921  
QY 980 CACACCTTAACACGGTGATATCCATCATATTAAGTTTTCAGCAACGGTTAG 1039  
Db 920 TTTCATACACACAGTTGACATCGATGTTATAGCGTGAATCGCTTGGGAATGACTGCG 861  
QY 1040 ATGATGTCACAGCTCTCACTGAAGGTATAA 1069  
Db 860 TCGTGGCATTTCTTTCGCTAAAGGTCAA 831  
RESULT 24  
US/08/622  
; INFORMATION FOR SEQ ID NO: 11:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1347 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA to mRNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 53..1168  
; SEQUENCE DESCRIPTION: SEQ ID NO: 11:  
US/08/622,679D-11  
Query Match 3.1%; Score 34.8; DB 3; Length 1347;  
Best Local Similarity 48.5%; Pred. No. 4.9;  
Matches 96; Conservative 0; Mismatches 102; Indels 0; Gaps 0;  
QY 680 TTATTGAGCAGATCAAAATCGAAGTATCCAAACCTGCAACATGTTATTGACGCTGTGGAA 739  
Db 978 TTCTGGGCAAAATTCCTGAGCAACAGGAACCAATATTATTTTTGGGATGATTTCATC 1037  
QY 740 GCGAAGATAGTATCCCGAGGCGCTATAAAGTTCACAGCAGATAGTCTACCTGCCACATTAT 799

Db 1038 CCAAGTGAAGTAGTCAATGAAGCATATTCAAGATTAGCATATTCTGCGTTATCTCTCAATAC 1097  
 QY 800 TAGAAGTGGTTCATGACCATTCAGAGCATTCCTGAAGAAATCAGAAAGATAATGTTA 859  
 Db 1098 TTGATGCTGATCTCTTGGCCATTCGCGCTTAACAGGCAAAAATGTCATGATAAAGTGA 1157  
 QY 860 AAATTGATATTACTTTGT 877  
 Db 1158 AGATAACAATAGACTGTAT 1175

RESULT 25  
 US-09-165-922A-11  
 ; Sequence 11, Application US/09165922A  
 ; Patent No. 6348641  
 ; GENERAL INFORMATION:  
 ; APPLICANT: STILES, JOHN I.  
 ; MOISYADI, ISTEF  
 ; NEUPANE, KABI R.  
 ; TITLE OF INVENTION: PURIFIED PROTEINS,  
 ; RECOMBINANT DNA SEQUENCES AND PROCESSES FOR  
 ; PRODUCING CAFFEINE FREE BEVERAGES  
 ; NUMBER OF SEQUENCES: 11  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: JONES, DAY, REAVIS & FOGUE  
 ; STREET: NORTH POINT, 901 LAKESIDE AVENUE  
 ; CITY: CLEVELAND  
 ; STATE: OHIO  
 ; COUNTRY: USA  
 ; ZIP: 44114  
 ; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb  
 ; COMPUTER READABLE FORM:  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: MS-DOS V. 5.1  
 ; SOFTWARE: WordPerfect for Windows V. 6.1  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/165,922A  
 ; FILING DATE: 02-Oct-1998  
 ; CLASSIFICATION: 800  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: GRIFFITH, CALVIN P.  
 ; REGISTRATION NUMBER: 34,831  
 ; REFERENCE/DOCKET NUMBER: 265036600003  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (216) 586-7050  
 ; TELEFAX: (216) 579-0212  
 ; INFORMATION FOR SEQ ID NO: 11:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1347 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: cdna to mRNA  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 53..1168  
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 11:

Query Match 3.1%; Score 34.8; DB 4; Length 1347;  
 Best Local Similarity 48.5%; Pred. No. 4.9;  
 Matches 96; Conservative 0; Mismatches 102; Indels 0; Gaps 0;  
 QY 680 TTATTGACGAGATCAAAATCGAAGTATCCAAACCTGCAACATGTTATTGACGCTGTGGAA 739  
 Db 978 TTCTGGGCAAAATCCCTGCGAGCAACAGGAACCAATATTATTTTGGGATGATTCCATC 1037  
 QY 740 GCGAAGTAGTATCCCGAGCCCTATAAAGTCACGACGATAGTCTACCTGCCACATTAT 799  
 Db 1038 CCAAGTGAAGTAGTCAATGAAGCATATTCAAGATTAGCATATTCTGCGTTATCTCTCAATAC 1097  
 QY 800 TAGAAGTGGTTCATGACCATTCGAAAGCATTCCTGAAGAAATCAGAAAGATAATGTTA 859

Db 1098 TTGATGCTGATCTCTTGGCCATTCGCGCTTAACAGGCAAAAATGTCATGATAAAGTGA 1157  
 QY 860 AAATTGATATTACTTTGT 877  
 Db 1158 AGATAACAATAGACTGTAT 1175

RESULT 26  
 US-09-601-198-131/c  
 ; Sequence 131, Application US/09601198  
 ; Patent No. 6531583  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Cassell, Gail H.  
 ; APPLICANT: Chen, Ellison Y.  
 ; APPLICANT: Glass, Jennifer S.  
 ; APPLICANT: Heiner, John I.  
 ; APPLICANT: Lefkowitz, Elliot  
 ; TITLE OF INVENTION: NUCLEIC ACID PROBES AND METHOD FOR DETECTING UREA PLASMA  
 ; FILE REFERENCE: UAB-13452/22  
 ; CURRENT APPLICATION NUMBER: US/09/601,198  
 ; PRIOR FILING DATE: 2000-12-08  
 ; PRIOR APPLICATION NUMBER: 60/073,189  
 ; PRIOR FILING DATE: 1998-01-30  
 ; NUMBER OF SEQ ID NOS: 181  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 131  
 ; LENGTH: 1374  
 ; TYPE: DNA  
 ; ORGANISM: Ureaplasma urealyticum  
 ; US-09-601-198-131

Query Match 3.1%; Score 34.8; DB 4; Length 1374;  
 Best Local Similarity 43.3%; Pred. No. 4.9;  
 Matches 162; Conservative 0; Mismatches 212; Indels 0; Gaps 0;  
 QY 732 TGTGGGAACGAGATAGTATCCCGAGCCCTATAAAGTCACAGCAGATAGTCTACCTGC 791  
 Db 976 TGTGTGGATTATATATAAATAATCAAGTTTGTCTAACTTTAAGCATGGTATTATGT 917  
 QY 792 CACATTATTAGAAAGTGGTTCCCAATGACCATTCGAAGCATTCCTGAAGAAATCAGAAAGA 851  
 Db 916 TAATTATGATGAATTAATAAATAATTAATGCGATTATAATGTTTCATGAAGATTAAAGACCA 857  
 QY 852 TAATGTTAAATTTGATATTACTTTGTTGTTATCGTGCATCTGGTCAAGAAATTCATTGGG 911  
 Db 856 TTATATATCTAAAAATAACCTGTTGTTAATAATAATTTTGAATTTAAACCATCACAATAT 797  
 QY 912 TGCACAAAGATTTCCTGCTAGTCCAGAAATATCATGAAGCCAGATTAAATTCGTTAAGTT 971  
 Db 796 TAATAAGTGAAGTAGCATTAACGCATCTCGTAAATTTAGCAACATTAAATGGCTTTATTTT 737  
 QY 972 TATAATCCACACCTTTAACAACGGTGATATCCATCATATTAATATAAAGTTTTCAGCAA 1031  
 Db 736 AAATGATTGTACACTACCAATGATTAAACAAGATAAAGCAATTAAGTTTATGATAA 677  
 QY 1032 CGGCTTAGATGATGTCCTCCAGCTCTCCTCAAGGATATAAAGAGGTAAACAAATAATGT 1091  
 Db 676 TGTATTGCTAAATTTGTTGATATTTCATGACGTTAAATAATTTAGCATGATATGACAAAA 617  
 QY 1092 TAAGTATGTTGCCA 1105  
 Db 616 TAATAACAATGCAA 603

RESULT 27  
 US-09-173-053-3  
 ; Sequence 3, Application US/09173053  
 ; Patent No. 6451769  
 ; GENERAL INFORMATION:  
 ; APPLICANT: HUEBNER, Robert C.

```

; APPLICANT: NORMAN, Jon A.
; APPLICANT: LIANG, Xiaowu
; APPLICANT: CARNER, Kristin R.
; APPLICANT: BARBOUR, Alan G.
; APPLICANT: LUKE, Catherine J.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR ADMINISTERING BORRELIA DNA
; FILE REFERENCE: 454312-2440.1
; CURRENT APPLICATION NUMBER: US/09/173,053
; CURRENT FILING DATE: 1998-10-15
; PRIOR APPLICATION NUMBER: 08/663,998
; PRIOR FILING DATE: 1996-06-14
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 4016
; TYPE: DNA
; ORGANISM: Borrelia burgdorferi
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (6)..(29)
; OTHER INFORMATION: N stands for A or G or C or T
; NAME/KEY: misc_feature
; LOCATION: (1078)..(1085)
; OTHER INFORMATION: N stands for A or G or C or T
US-09-173-053-3

```

```

RESULT 20
US-08-781-891-209
; Sequence 209, Application US/08781891
; Patent No. 6090620
; GENERAL INFORMATION:
; APPLICANT: Fu, Ying-Hui
; APPLICANT: Yu, Chang-En
; APPLICANT: Oshima, Junko
; APPLICANT: Mulligan, John T.
; APPLICANT: Schellenberg, Gerald D.
; TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
; TITLE OF INVENTION: WERNER'S S SYNDROME
; NUMBER OF SEQUENCES: 209
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington

```

```

RESULT 29
US-09-618-166-209
; Sequence 209, Application US/09618166
; Patent No. 6583112
; GENERAL INFORMATION:
; APPLICANT: Fu, Ying-Hui
; Yu, Chang-En
; Oshima, Junko
; Mulligan, John T.
; Schellenberg, Gerald D.
; TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO TO
; WERNER'S SYNDROME
; NUMBER OF SEQUENCES: 209
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed Intellectual Property Law Group
; STREET: 701 Fifth Avenue, Suite 6300
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/618,166  
FILING DATE: 17-Jul-2000  
CLASSIFICATION: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Mcmasters, David D.  
REGISTRATION NUMBER: 33,963  
REFERENCE/DOCKET NUMBER: 240052.419C1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 209:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 51259 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 209:  
US-09-618-166-209  
Query Match 3.1%; Score 34.8; DB 4; Length 51259;  
Best Local Similarity 49.2%; Pred. No. 21;  
Matches 124; Conservative 0; Mismatches 122; Indels 6; Gaps 1;  
QY 187 TGTGACATTCCTGGTACAGTGTGCAACATTTGGACCAATCTAGTACTGACTTGAAGTT 246  
Db 14400 TGGGATGTTATTTGTACAGTAGTCTGACATTTAACTTAATCAGATTTGTCTATTTTAGGTA 14459  
QY 247 GGAGATACCGTTTCGGTTTGTTCACGGTGTCTCCCAACACAGATCCTAAATGGTGCA 306  
Db 14460 AATGTTACATTTTCTTAAGTAGTCCGGTCTATAACAGAAATAGCAAGCATCTTCA 14519  
QY 307 TTGTCTGAATATGACGGGTTTATCCACCTTTGTTTAA-----CAAGAGTAACTTAACT 360  
Db 14520 TGGGGTCCCTTCCAGCGGTACTTGTGATTGTCTTTTAACTTTGGGAATGAGACTTGAAT 14579  
QY 361 CACTCAACTGCTGATGAATTTCTGAAGGCCCTGTGAGAACTTCGAACTCTGCTGCATCA 420  
Db 14580 GGCAGATGCCTAAATGAATCTCTACAGGACCTTGGGAAGCCCTTGAACCTTTTGCATTCA 14639  
QY 421 TTGCCAGTTTCG 432  
Db 14640 GAGTGAATTTTG 14651

RESULT 30  
US-08-360-606B-29  
Sequence 29 Application US/08360606B  
Patent No. 5919617  
GENERAL INFORMATION:  
APPLICANT: Jnanendra K. Bhattacharjee  
APPLICANT: Richard C. Garrod  
APPLICANT: Paul L. Skatrud  
APPLICANT: Robert P. Peery  
TITLE OF INVENTION: Methods and Reagents for  
TITLE OF INVENTION: Detecting Fungal Pathogens in a  
TITLE OF INVENTION: Biological Sample  
NUMBER OF SEQUENCES: 33  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff  
STREET: 300 S. Wacker Drive Suite 3200  
CITY: Chicago  
STATE: Illinois  
COUNTRY: U.S.A.  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: MS Word 7.0

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/360,606B  
FILING DATE: December 21, 1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Berghoff, Paul H.  
REGISTRATION NUMBER: 30,243  
REFERENCE/DOCKET NUMBER: 94,319  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312)913-0001  
TELEFAX: (312)913-0002  
INFORMATION FOR SEQ ID NO: 29:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1856 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Saccharomyces cerevisiae  
US-08-360-606B-29  
Query Match 3.1%; Score 34.4; DB 2; Length 1856;  
Best Local Similarity 48.5%; Pred. No. 7.1;  
Matches 95; Conservative 0; Mismatches 101; Indels 0; Gaps 0;  
QY 22 AAAGCGTTCATCATTTGAAGGTGACAAAGCTGTTGTTAAACACAGATGCTCAGTTCAGAA 81  
Db 973 AAAGATGTTAAATTTGAATTGAAAAAGCTTTAACTAAAAATGGGGGTCAATATCTCTAAA 1032  
QY 82 TTAAGAGGGGTACAGCCTTTGGTGAAGTTGAGGCTTTGCTGTTAAACCACTGATTGG 141  
Db 1033 TGTCTGTTTATTGGTGGCTTTGGGTAGATGCTGGTCCCATTTATTATTAATAAAA 1092  
QY 142 AAGCATATTGCTTATAAGATTGGTCCAGAGGTTCAATTTCTAGGATGTGACATTTGCTGTT 201  
Db 1093 ATTGGTATCCCTGATGATAATATTGCTAAATGGGATATGGCTGAAACTGCTAAAGGTGTT 1152  
QY 202 ACAGTTGTCAAACTTG 217  
Db 1153 CCATTCCAAGAAATTG 1168  
Search completed: January 10, 2004, 08:41:27  
Job time : 112 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 10, 2004, 07:53:26 ; Search time 448 Seconds  
(without alignments)  
8627.273 Million cell updates/sec

Title: US-10-081-644-1

Perfect score: 1113

Sequence: 1 atgtcagttcccaaccactca.....agtatgttccaggtataa 1113

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2276164 seqs, 1736306516 residues

Total number of hits satisfying chosen parameters: 4552328

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 90 summaries

Database : Published Applications\_NA:\*

- 1: /cgn2\_6/prodata/2/pubpna/PCT\_NEW\_PUB.seq.\*
- 2: /cgn2\_6/prodata/2/pubpna/PCT\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/prodata/2/pubpna/US06\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/prodata/2/pubpna/US06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/prodata/2/pubpna/US07\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/prodata/2/pubpna/US07\_PUBCOMB.seq.\*
- 7: /cgn2\_6/prodata/2/pubpna/US08\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/prodata/2/pubpna/US08\_PUBCOMB.seq.\*
- 9: /cgn2\_6/prodata/2/pubpna/US09A\_PUBCOMB.seq.\*
- 10: /cgn2\_6/prodata/2/pubpna/US09B\_PUBCOMB.seq.\*
- 11: /cgn2\_6/prodata/2/pubpna/US09C\_PUBCOMB.seq.\*
- 12: /cgn2\_6/prodata/2/pubpna/US09\_NEW\_PUB.seq.\*
- 13: /cgn2\_6/prodata/2/pubpna/US09\_NEW\_PUB.seq.\*
- 14: /cgn2\_6/prodata/2/pubpna/US10A\_PUBCOMB.seq.\*
- 15: /cgn2\_6/prodata/2/pubpna/US10B\_PUBCOMB.seq.\*
- 16: /cgn2\_6/prodata/2/pubpna/US10\_NEW\_PUB.seq.\*
- 17: /cgn2\_6/prodata/2/pubpna/US60\_NEW\_PUB.seq.\*
- 18: /cgn2\_6/prodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1113	100.0	1113	14	US-10-081-644-1
2	502.2	45.1	509	14	US-10-081-644-1
3	359	32.3	1134	14	US-10-081-644-5
4	357.8	32.1	1145	14	US-10-081-644-3
5	345.2	31.0	1122	14	US-10-081-644-7
6	45.4	4.1	23439	8	US-08-781-986A-38
7	41	3.7	572	13	US-10-029-386-25419
8	41	3.7	3673778	13	US-10-312-841-1
9	40.6	3.6	3673778	13	US-10-312-841-2
10	40.2	3.6	439	9	US-09-864-761-20174
11	40	3.6	1236	15	US-10-077-584-3
12	39	3.5	1168	12	US-10-292-798-1825
13	38.8	3.5	332	10	US-09-878-574-1620
14	38	3.4	506	13	US-10-029-386-20619
15	38	3.4	1168	13	US-10-017-161-2179

3.4	10668	13	US-09-814-353-20029	Sequence 20029, A
3.4	640681	10	US-09-790-988-1	Sequence 1, Appl
3.4	573	13	US-10-029-386-25433	Sequence 25433, A
3.4	835	13	US-10-029-386-22786	Sequence 22786, A
3.4	1200	12	US-10-369-493-42193	Sequence 42193, A
3.3	7536	13	US-10-252-157-159	Sequence 159, App
3.3	7772	15	US-10-198-846-14020	Sequence 14020, A
3.3	8772	9	US-09-788-711A-3	Sequence 3, Appl
3.3	8871	9	US-09-788-711A-1	Sequence 1, Appl
3.3	10531	15	US-10-225-567A-523	Sequence 523, App
3.3	24370	12	US-10-292-798-931	Sequence 931, App
3.3	198285	10	US-09-880-107-3814	Sequence 3814, App
3.3	441	9	US-09-864-761-3403	Sequence 3403, App
3.3	3150	10	US-09-938-842A-243	Sequence 243, App
3.3	969	9	US-09-815-243-6606	Sequence 6606, App
3.3	6253	10	US-09-070-927A-160	Sequence 160, App
3.3	4444	12	US-10-369-493-30105	Sequence 30105, A
3.3	803	13	US-10-027-632-126552	Sequence 126552, A
3.3	803	13	US-10-027-632-171708	Sequence 171708, A
3.3	803	13	US-10-027-632-171709	Sequence 171709, A
3.3	803	14	US-10-027-632-126552	Sequence 126552, A
3.3	803	14	US-10-027-632-171708	Sequence 171708, A
3.3	803	14	US-10-027-632-171709	Sequence 171709, A
3.3	3255	13	US-10-349-680-32	Sequence 32, Appl
3.3	580073	13	US-10-205-220-1	Sequence 1, Appl
3.2	312	9	US-09-815-242-7011	Sequence 7011, App
3.2	9095	13	US-10-240-453-322	Sequence 322, App
3.2	1830121	15	US-10-329-960-1	Sequence 1, Appl
3.2	586	13	US-10-027-632-11312	Sequence 11312, A
3.2	586	13	US-10-027-632-11312	Sequence 11312, A
3.2	692	13	US-10-027-632-104677	Sequence 104677, A
3.2	692	14	US-10-027-632-104677	Sequence 104677, A
3.2	10806	13	US-10-311-455-10	Sequence 10, Appl
3.2	429	9	US-09-815-242-7144	Sequence 7144, App
3.2	2421	9	US-09-815-242-3860	Sequence 3860, App
3.2	2424	9	US-09-815-242-6622	Sequence 6622, App
3.2	3312	13	US-10-349-680-150	Sequence 150, App
3.2	10929	15	US-10-032-393-7	Sequence 7, Appl
3.2	520	9	US-09-864-761-8552	Sequence 8552, App
3.2	1803	10	US-09-887-576-812	Sequence 812, App
3.2	1959	9	US-09-864-761-4012	Sequence 4012, App
3.2	186510	13	US-10-043-715-1	Sequence 1, Appl
3.2	405	10	US-09-960-352-13631	Sequence 13631, A
3.2	538	13	US-10-029-386-20406	Sequence 20406, A
3.2	2000	10	US-09-938-842A-3404	Sequence 3404, App
3.2	6437	13	US-10-311-455-1234	Sequence 1234, App
3.2	7309	13	US-10-311-455-1789	Sequence 1789, App
3.2	640681	10	US-09-790-988-1	Sequence 1, Appl
3.1	611	13	US-10-027-632-216679	Sequence 216679, A
3.1	611	14	US-10-027-632-216679	Sequence 216679, A
3.1	1482	10	US-09-070-927A-842	Sequence 842, App
3.1	3616	13	US-10-070-923-4	Sequence 4, Appl
3.1	5513	9	US-09-800-065-3	Sequence 3, Appl
3.1	13431	13	US-10-070-923-1	Sequence 1, Appl
3.1	47309	12	US-10-085-117-79	Sequence 79, Appl
3.1	360	10	US-09-736-457-1408	Sequence 1408, App
3.1	360	10	US-09-902-941-1408	Sequence 1408, App
3.1	360	10	US-09-849-626-1408	Sequence 1408, App
3.1	360	13	US-10-113-872-1408	Sequence 1408, App
3.1	456	9	US-09-864-761-4249	Sequence 4249, App
3.1	1374	13	US-10-349-680-53	Sequence 53, Appl
3.1	3931	15	US-10-006-780-1	Sequence 1, Appl
3.1	4330	12	US-10-369-493-25222	Sequence 25222, A
3.1	477	9	US-09-864-761-5436	Sequence 5436, App
3.1	503	11	US-09-764-891-9331	Sequence 9331, App
3.1	503	15	US-10-091-572-878	Sequence 878, App
3.1	945	12	US-10-369-493-42143	Sequence 42143, A
3.1	1080	10	US-09-938-842A-231	Sequence 231, App
3.1	1283	10	US-09-887-576-601	Sequence 601, App
3.1	17858	12	US-10-242-355-1233	Sequence 1233, App
3.1	17858	13	US-10-080-254-147	Sequence 147, App
3.1	142299	11	US-09-911-077A-14	Sequence 14, Appl

c 89 34.4 3.1 418 11 US-09-991-936-213 Sequence 213, App  
c 90 34.4 3.1 553 13 US-10-027-632-277990 Sequence 277990,  
  
ALIGNMENTS  
  
RESULT 1  
US-10-081-644-1  
; Sequence 1, Application US/10081644  
; Publication No. US20020192782A1  
; GENERAL INFORMATION:  
; APPLICANT: Yamamoto, Hiroaki  
; APPLICANT: Kimoto, No. US20020192782A1hiro  
; TITLE OF INVENTION: NOVEL ENONE REDUCTASES, METHODS FOR  
; TITLE OF INVENTION: PRODUCING SAME, AND METHODS FOR SELECTIVELY REDUCING A  
; TITLE OF INVENTION: CARBON-CARBON DOUBLE BOND OF AN ALPHA, BETA-UNSATURATED KETONE  
; TITLE OF INVENTION: USING THE REDUCTASES  
; FILE REFERENCE: 06501-100001  
; CURRENT APPLICATION NUMBER: US/10/081,644  
; PRIOR FILING DATE: 2002-02-21  
; PRIOR APPLICATION NUMBER: JP 2001-49363  
; PRIOR FILING DATE: 2001-02-23  
; NUMBER OF SEQ ID NOS: 29  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 1113  
; TYPE: DNA  
; ORGANISM: Kluyveromyces lactis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)...(1110)  
US-10-081-644-1  
  
Query Match 100.0%; Score 1113; DB 14; Length 1113;  
Best Local Similarity 100.0%; Pred. No. 7.6e-284;  
Matches 1113; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 ATCTGAGTCCCAACCACTCAAAAGCGCTCATCATTTGAAGTGCACAAAGCTGTGTAA 60  
Db 1 ATCTGAGTCCCAACCACTCAAAAGCGCTCATCATTTGAAGTGCACAAAGCTGTGTAA 60  
  
QY 61 ACAGATGCTCTCAGTTCAGATTAAGGAGGGTACAGCCTTGGTGAAGTGTGAGGCTGTT 120  
Db 61 ACAGATGCTCTCAGTTCAGATTAAGGAGGGTACAGCCTTGGTGAAGTGTGAGGCTGTT 120  
  
QY 121 GCTGGTAAACCACTGATTCGAGATTAAGTGTCTTATAAGATTTGGTCCAGAAGTTCAAAT 180  
Db 121 GCTGGTAAACCACTGATTCGAGATTAAGTGTCTTATAAGATTTGGTCCAGAAGTTCAAAT 180  
  
QY 181 CTAGGATGTGACATTTGCTGTGATGATGATGATGATGATGATGATGATGATGATGATG 240  
Db 181 CTAGGATGTGACATTTGCTGTGATGATGATGATGATGATGATGATGATGATGATGATG 240  
  
QY 241 AAGTGTGGAGATACCGGTTTTCGGTTCAGGTCCTTCCAGGTCCTTCCAAACAGATCCTAAAT 300  
Db 241 AAGTGTGGAGATACCGGTTTTCGGTTCAGGTCCTTCCAGGTCCTTCCAAACAGATCCTAAAT 300  
  
QY 301 GGTGCATTTGCTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 360  
Db 301 GGTGCATTTGCTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 360  
  
QY 361 CACTCACTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 420  
Db 361 CACTCACTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 420  
  
QY 421 TTGCCAGTTTCGTTGACAACTGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 480  
Db 421 TTGCCAGTTTCGTTGACAACTGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 480  
  
QY 481 GAATGGCACCCTACTACCCCGCAACATACATCATCCATTTATGATTTGGGGTGTGTCTACA 540  
Db 481 GAATGGCACCCTACTACCCCGCAACATACATCATCCATTTATGATTTGGGGTGTGTCTACA 540

QY 541 GCAGTGGTCAACAACCTAATCAAGTTGCCAAACATATCAATGCTTTACTAAGATTGTA 600  
Db 541 GCAGTGGTCAACAACCTAATCAAGTTGCCAAACATATCAATGCTTTACTAAGATTGTA 600  
  
QY 601 ACTGTTGCTTTCTAAAAAGCATGAAAAGCTTTTAAAGTCTTTATGGTCTGATGATGCTTT 660  
Db 601 ACTGTTGCTTTCTAAAAAGCATGAAAAGCTTTTAAAGTCTTTATGGTCTGATGATGCTTT 660  
  
QY 661 GACTATCATGATGACAGGCGTTTATGAGCAGATCAAAATCGAAGTATCCAAACCTGCAACAT 720  
Db 661 GACTATCATGATGACAGGCGTTTATGAGCAGATCAAAATCGAAGTATCCAAACCTGCAACAT 720  
  
QY 721 GTTATTGACGCTGTGGGAAGCAAGATAGTATCCCGAGCGCTTATAAAGTCAACAGCAT 780  
Db 721 GTTATTGACGCTGTGGGAAGCAAGATAGTATCCCGAGCGCTTATAAAGTCAACAGCAT 780  
  
QY 781 AGTCTACCTGCCACATTTATTAGAAGTGGTTCCAAATGACCATTTGAAAGCATTTCTGAAAG 840  
Db 781 AGTCTACCTGCCACATTTATTAGAAGTGGTTCCAAATGACCATTTGAAAGCATTTCTGAAAG 840  
  
QY 841 ATCAGAAAAGATAAGTTTAAATTTGATATTTCTTTGTTGTTGTTGTTGTTGTTGTTGTT 900  
Db 841 ATCAGAAAAGATAAGTTTAAATTTGATATTTCTTTGTTGTTGTTGTTGTTGTTGTTGTT 900  
  
QY 901 ATTCTATTGGGTGCAACAAGATTTCTGCTAGTCCAGATATATCATGAAGCCACAGTTAA 960  
Db 901 ATTCTATTGGGTGCAACAAGATTTCTGCTAGTCCAGATATATCATGAAGCCACAGTTAA 960  
  
QY 961 TTCGTTAAGTTTATAAATCCACACCTTAAACAAGGTGATATCCATCATATGATATATA 1020  
Db 961 TTCGTTAAGTTTATAAATCCACACCTTAAACAAGGTGATATCCATCATATGATATATA 1020  
  
QY 1021 GTTTTCAGCAACGGCTTAGATGATGATGATGATGATGATGATGATGATGATGATGATG 1080  
Db 1021 GTTTTCAGCAACGGCTTAGATGATGATGATGATGATGATGATGATGATGATGATGATG 1080  
  
QY 1081 AACAAAAATGTTAAAGTATGTTGCCAGGTTATA 1113  
Db 1081 AACAAAAATGTTAAAGTATGTTGCCAGGTTATA 1113  
  
RESULT 2  
US-10-081-644-14  
; Sequence 14, Application US/10081644  
; Publication No. US20020192782A1  
; GENERAL INFORMATION:  
; APPLICANT: Yamamoto, Hiroaki  
; APPLICANT: Kimoto, No. US20020192782A1hiro  
; TITLE OF INVENTION: NOVEL ENONE REDUCTASES, METHODS FOR  
; TITLE OF INVENTION: PRODUCING SAME, AND METHODS FOR SELECTIVELY REDUCING A  
; TITLE OF INVENTION: CARBON-CARBON DOUBLE BOND OF AN ALPHA, BETA-UNSATURATED KETONE  
; TITLE OF INVENTION: USING THE REDUCTASES  
; FILE REFERENCE: 06501-100001  
; CURRENT APPLICATION NUMBER: US/10/081,644  
; PRIOR FILING DATE: 2002-02-21  
; PRIOR APPLICATION NUMBER: JP 2001-49363  
; PRIOR FILING DATE: 2001-02-23  
; NUMBER OF SEQ ID NOS: 29  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 14  
; LENGTH: 509  
; TYPE: DNA  
; ORGANISM: Kluyveromyces lactis  
US-10-081-644-14  
  
Query Match 45.1%; Score 502.2; DB 14; Length 509;  
Best Local Similarity 98.4%; Pred. No. 1.9e-122;  
Matches 501; Conservative 5; Mismatches 3; Indels 0; Gaps 0;  
  
QY 160 ATTGGTCCAGAGGTTCAATTTAGATGTGACATTTGGTACAGTTGTCAAACCTTGA 219  
Db 1 ATTGGTCCWAGGGVTCWATTTCTAGATGTGACATTTGGTACAGTTGTCAAACCTTGA 60

Qy 220 CCAAAATGCTAGTACTGACTTGAAGGTTGGAGATACCGGTTTCGGTTTGGTTTCAACGGTGCT 279  
Db 61 CCAAAATGCTAGTACTGACTTGAAGGTTGGAGATACCGGTTTCGGTTTGGTTTCAACGGTGCT 120  
Qy 280 TCCCAACAGATCCTAAAATGGTGCAATTCCTGATATGCCAGGGTTTATCCACCTTTG 339  
Db 121 TCCCAACAGATCCTAAAATGGTGCAATTCCTGATATGCCAGGGTTTATCCACCTTTG 180  
Qy 340 TTTTACAAGAGTAACCTAACTCACTCACTGCTGATGAATTTCTGAAGGCCCTGTGAAG 399  
Db 181 TTTTACAAGAGTAACCTAACTCACTCACTGCTGATGAATTTCTGAAGGCCCTGTGAAG 240  
Qy 400 AACTTCGAATCTGCTGCATCATTCGCCAGTTTCGTTGACAACTGCTGGTGTAGTTGTGT 459  
Db 241 AACTTCGAATCTGCTGCATCATTCGCCAGTTTCGTTGACAACTGCTGGTGTAGTTGTGT 300  
Qy 460 CATCACTGGGCTCAAAAATGGATGGCACCCTATCCCGGCCCAACATCACTCACTCAATTA 519  
Db 301 CATCACTGGGCTCAAAAATGGATGGCACCCTATCCCGGCCCAACATCACTCACTCAATTA 360  
Qy 520 TTGATTTGGGGTGGTGTACAGAGTGGTCAACAACCTAACTCCAAAGTTGCCAAACATATC 579  
Db 361 TTGATTTGGGGTGGTGTACAGAGTGGTCAACAACCTAACTCCAAAGTTGCCAAACATATC 420  
Qy 580 AATGCTTATACCTAAGATTGTAACCTGTTGCTTCTAAAAGAGCATGAAAAGCTTTTAAAGTCT 639  
Db 421 AATGCTTATACCTAAGATTGTAACCTGTTGCTTCTAAAAGAGCATGAAAAGCTTTTAAAGTCT 480  
Qy 640 TATGGTGTGATGATGCTTTGACTATCA 668  
Db 481 TATGGTGTGATGATGCTTTGACTATCA 509

## RESULT 3

US-10-081-644-5  
; Sequence 5, Application US/10081644  
; Publication No. US20020192782A1  
; GENERAL INFORMATION:  
; APPLICANT: Yamamoto, Hiroaki  
; APPLICANT: Kimoto, No. US20020192782A1hiro  
; TITLE OF INVENTION: NOVEL ENGINE REDUCTASES, METHODS FOR  
; TITLE OF INVENTION: PRODUCING SAME, AND METHODS FOR SELECTIVELY REDUCING A  
; TITLE OF INVENTION: CARBON-CARBON DOUBLE BOND OF AN ALPHA, BETA-UNSATURATED  
; TITLE OF INVENTION: USING THE REDUCTASES  
; FILE REFERENCE: 06501-100001  
; CURRENT APPLICATION NUMBER: US/10/081,644  
; CURRENT FILING DATE: 2002-02-21  
; PRIOR APPLICATION NUMBER: JP 2001-49363  
; PRIOR FILING DATE: 2001-02-23  
; NUMBER OF SEQ ID NOS: 29  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 5  
; LENGTH: 1134  
; TYPE: DNA  
; ORGANISM: Saccharomyces cerevisiae  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)...(1131)  
US-10-081-644-5

Query Match 32.3%; Score 359; DB 14; Length 1134;  
Best Local Similarity 58.9%; Pred. No. 2, 4e-84;  
Matches 658; Conservative 0; Mismatches 450; Indels 9; Gaps 2;

Qy 5 CAGTTCCCAACCACTCAAAAGCCGTCATCTTGAAGGTGACAAAGCTGTGTTTAAACAG 64  
Db 14 CAATTCAGAAACCAATGAAGGCTGCTGCTCATTTGAAGACGGTAAAGCGGTTGTTAAAGAG 73  
Qy 65 ATGTCCTAGTTCCAGAAATTAAGAGGGTACAGCCCTTGGTGAAGTTGAGGCTGTGCTG 124  
Db 74 GCATTCCTATTCCTGAATTTGAAGAGGATTCGTTATTTGAAGACACTCGCTGTGCTG 133

## RESULT 4

US-10-081-644-3  
; Sequence 3, Application US/10081644  
; Publication No. US20020192782A1

Qy 125 GTAAACCAACTGATTTGAAGCATATATGCTTATAGATTGCTCCAGAAAGGTTCAATTCCTAG 184  
Db 134 GTAAACCAACTGATTTGGGCACACATTTGACTACAAGATCGGGCTTCAAGGATCTATTCTGG 193  
Qy 185 GATGTGACATTCGCTGGTACAGTTTGTCAAACTTGGACCAAAATGCTAGTACT--GACTTGA 241  
Db 194 GATGTGATGCTGCTGGCCAAATTTGTCAAAATTTGGGCCAGCTGTCAATCTTAAAGACTTTT 253  
Qy 242 AGGTTGAGATACCGGTTTCGGTTTGTTCACCGGTTTCCCAAAACAGATCTTAAAAAATG 301  
Db 254 CTATCGGTGATATATTTATGGGTTTCACTACCGATCTTCGGTAAGGTTTCTTCCCAATG 313  
Qy 302 GTGCAATTTGCTGAATATGCGAGGGTTTATCCACTTTGTTTAAACA-----GAGTAACT 355  
Db 314 GTGCTTTTGTGTAATATTTCTGCTATTTCAACTGTGGTGGCTTACAAATCACCCTCAATGAAC 373  
Qy 356 TAACTCACTCAACTGCTGATGAATTTCTGAAGGCCCTGTGAAGAACTTCGAATCTGCTG 415  
Db 374 TCAAAATTTTGGGTGAGGATGTTCTACCTCCCGGCCCTGTCAAGTCTTTGGAGGTTGAG 433  
Qy 416 CATCAATGCCAGTTTCGTTGACAACTGCTGTTAGTTTGTGTGTCATCACTTTGGGCTCAA 475  
Db 434 CCACATATCCAGTGTCACTGACCAAGCCGCTTGGTGTGACCTATAACTTTGGGCTTGG 493  
Qy 476 AAATGGAATGGCACCCATCTACCCCGCAACATCTCATCCATTTATGATTTGGGGTGGT 535  
Db 494 ACCTGAAGTGGAGCCATCAACCCCAAAAGAAAGGCCCCATCTTATTTATGGGGCGGTG 553  
Qy 536 CTACAGCAGTGGGTCAACAACCTAATCCAGTTGGCAACATATCAATGCTTTATCTAAGA 595  
Db 554 CAACCTGCAGTAGGTGAGTCTGCTCATCCAAATTTAGCCAAATTAATGAAATGGGTTTCA 613  
Qy 596 TTGTAACCTGTTGCTTCTAAAAGCATGAAAAGCTTTTAAAAGTCTTTATGGTCTGATGAT 655  
Db 614 TCATTTGTTGGCTTCTCGGAAGCAAGAACTTTTGAAGAATATGTTGCTGCTGATGAT 673  
Qy 656 TCTTTGACTATCATGATGACAGGCTTATTTGAGCAGATCAAAATCGAAGTATCCAAACCTGC 715  
Db 674 TATTTGATTTATCATGATATTTGACGTGTGAGCAAAATTTAAACACAACTATATCT 733  
Qy 716 AACATGTTATGACGCTGTGGGAAGCGAAGATAGTATCCCGAGGCTTATAAGTCAAG 775  
Db 734 CGTATTTAGTTCGACTGTGTCGCGAATCAAGATACGCTTCAACAAAGTGTACAAATGTGCGG 793  
Qy 776 CAGATGCTTACCTGCGCACATTTATGAAGTGGTTCCAATGACCATTTGAAAGCATTCCTG 835  
Db 794 CCGATAAACAGGATGCTACAAATTTGTTGATTTAAAATTTGACNAGAAAGAACGTCAAAA 853  
Qy 836 AAGAAATCAGAAAAGATAATGTTAAAATTTGATATTTACTTTTGTGTTATCGTCACTCGT 895  
Db 854 AAGAGAACAGGAGACAAAACGTTTACTATTGACATATAAAGCTATATTCAATAGGTGGCC 913  
Qy 896 AAGAAATTTATTTGGGTGCAACAGATTTCTGCTGCTAGTCAGAAATATCATGAAGCCACAG 955  
Db 914 ATGAAGTACCAATTTGGAACAATTTACTTTTACCAGCCGACTCAGAAGCTAGGAAAGCTCAA 973  
Qy 956 TTAATTTGTTAAAGTTTATAAATCCACACCTTAAACAGGTGATATCATCATATGAATA 1015  
Db 974 TAAATTTTATCAAAATTCATCAATCCAAAGATTAATGATGGACAAATTTGCCCATATTCAG 1033  
Qy 1016 TTAAGTTTTCAGCAACGGCTTAGATGATGTCGCCAGCTCTCACTGAAGGTTATAAAGAG 1075  
Db 1034 TAAGGTTCTATNAGAACGGGCTTTGTGATGTTCTCTCATATCTCTTAAGACATCAATATG 1093  
Qy 1076 GTAAAAACAAAAATGTTAAAGTATGTTGCCAGGTTATA 1112  
Db 1094 GTAAGAAGCTCGTGTGAAAAAAGCTCGTTCGCGGTATTTAAA 1130

	Query Match	31.0%;	Score 345.2;	DB 14;	Length 1122;
	Best Local Similarity	58.5%;	Pred. No. 1e-80;		
	Matches 643;	Conservative 0;	Mismatches 448;	Indels 9;	Gaps 2;
Qy	22	AAAGCGTCATCATTTGAAGGTGACAAAGCTGTGTTAAACACAGATGCTTCAGTTCCAGAA	81		
Db	10	AAAGCTGTCTCATTTGAAGACGGTAAAGCGGTTGTCAAGAGGGCGGTTCCCATTCCTGAA	69		
Qy	82	TTAAAGGAGGGGTACAGCCTTGGTCAAGGTTTGAAGGCTGTCTGCTGTAACCCAACTGATTGG	141		
Db	70	TTGGAGAAGGATTCTGATTGATTGAAGACACTCGCTGCTGCTGTAACCCGACTGATTGG	129		
Qy	142	AAGCATATTCTCTTATTAAGATTGTTCCAGAGGTTTCAATTCACGATGTGACTGCTGGT	201		
Db	130	GCACATATTGACTTACAAGGTTCGGGCTTCAAGGATCTATTCTGGGATGTGACGTCGTCGGC	189		



202 ACAGTTGTCAAATTTGGACCA---AATCTAGTACTGACTTTGAAGTTGGAGATACCGGT 258  
190 CAAATTTGCAAAATTTGGCCCGCGCGTGCATCTTAAGACCTTTCTATTGGTGAATATT 249  
259 TTGGTTTGTTCACGGTCTTCCCAACAGATCTTAAATAGTGATGCTTGTGAATAT 318  
250 TATGGTTTCATTACGGATCTTCCGTAAGTCTTCCATCGTGTCTTTGTGTAATAT 309  
319 GCCAGGGTTTATCACCTTTGTTTTACAA-----GAGTACTTAACCTCACTCACTGCT 372  
310 TCTGCTATTTCACTGTGGTGGTCTTACAAATACCCCAATGAACCTCAATTTTGGGTGAA 369  
373 GATGAATTTCTGAAGGCCCTGTGAAGAACTTCCGAATCTGCTGCATCATTCGCCAGTTTCG 432  
370 GATGTTCTACCTGCGGCCCTGTGAGTCTTTTGAAGGGCGAGCCACTATCCCACTGTC 429  
433 TTGACAACTGCTGTGTGTTAGTTGTGTCATCACTTTGGGCTCAAAATAGGAATGCCACCA 492  
430 CTGACCAAGCTGCTGTTGGTGTGACCTTAACTTTGGGCTTGAACCTGAAGTGGAGCCA 489  
493 TCTACCCCGCAACATCACTCACTTATTTGATTTGGGCTGTGTACAGCAGTGGGTCAA 552  
490 TCAACCCCAAGAAAGCGGCCCATCTTATATGGGCGGTGCACTGCAAGTAGGTGAG 549  
553 CAACATAATCCAAAGTTGCCAAACATATCAATGCTTATCTAAGATTGTAACTGTTGCTTCT 612  
550 TCGCTCATCAATTAGCCAAATAAATGAATGGCTTCCCAAGATCATTTGTTGGCTTCT 609  
613 AAAAGCATGAAAGCTTTTAAAGCTTTATGGTGTGATGATGCTTTTGAATCATGAT 672  
610 CGGAACACGCAAAACTGTTGAAAGATATGCTGCTGATCACTATTTGATTACCATGAT 669  
673 GCAGGCTTATTGACGAGATCAATCGAAGTATCCAACTGCAACATGTTATTGACGCT 732  
670 ATTGACGCTGTGAGAACAAATTAACACAAAGTACAAATATCTCGTATTTAGTGCAGTGT 729  
733 GTGGGAAGCAAGATAGTATCCCGAGGCTTATAAAGTACAGCAGATAGTCTACCTGCC 792  
730 GTCGGAATCAAAATACGCTTCAACAGTGTACAAATGTGGCGCGATTAACAGATGCT 789  
793 ACATTATTAGAAGTGTTCATGACCAATGAAAGCATCTCTGAAAGAAATCAGAAAGAT 852  
790 ACCGTTGTGCAATTAACCTAATTTGACAGAAAGAAACGTCAAAAGGAGAAATAGGAGCAA 849  
853 AATGTTAAATTTGATATTACTTTGTTGATCGTGCATCTGCTCAAGAAATCTATTGGT 912  
850 AATGTCACTATTGACAGAAACAGACTGTATTCAATAGCGCGCCATGAAGTACCATTGGT 909  
913 GCAACAAGATTTCTGCTAGTCCAGAAATCATGAAGCCACAGTTTAAATTCGTTAAGTTT 972  
910 GGCATTACTTTCCCTGCTGACCCAGAGAGGAGGCTGCCACCGAATTCGTCAGTTTC 969  
973 ATAAATCACAACCTTAAACAGCGGTGATATCATCATATGAAATATTAAGTTTTCAGCAAC 1032  
970 ATCAATCCAAAGATTAGTGTGGGCAAAATTCACCATATTTCCAGCAAGGCTCTATAAGAAC 1029  
1033 GCGTTAGATGATGCTCCAGCTCTCACTGAAGTATAAAGAGGTAAAGAAACAAATGTT 1092  
1030 GGGCTTTACGATGTTCTCGTATCTGGAAGACATTAATAATTCGGTAAGAACTCGTGTGAA 1089  
1093 AAGTATGTTGCCAGGTATA 1112  
1090 AACTCGTTGCGTATTAAA 1109

## RESULT 6

US-08-781-986A-38/c

; Sequence 38, Application US/08781986A  
; Publication No. US20030054436A1  
; GENERAL INFORMATION:  
; APPLICANT: Charles Kunsch  
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences

NUMBER OF SEQUENCES: 5255  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/781,986A  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Benson, Bob  
REGISTRATION NUMBER: 30,446  
REFERENCE/DOCKET NUMBER: PB248PP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 38:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 23439 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-781-986A-38

Query Match 4.1%; Score 45.4; DB 8; Length 23439;  
Best Local Similarity 45.4%; Pred. No. 1.3;  
Matches 163; Conservative 0; Mismatches 196; Indels 0; Gaps 0;

Qy 504 ACATACCTCATCCATTATTTGGGTGGTGTACAGCAGTGGTCAACAACTAATCCA 563  
Db 16078 ACAAGCTTTTAACTCTACGTTTATGCAGCGCAGGTAATGATATATGATTA 16019  
Qy 564 AGTTGCCAAACATATCAATGCTTATCTAAGATTGTAACCTGTTGCTTTCTAAAAGCATGA 623  
Db 16018 CATTCGTGATTACTAATTTCAAAATATATATTCACACAGATGAACCTTAAACACTTCC 15959  
Qy 624 AAAGCTTTTAAAGTCTTTATGGTGTGATGATGCTTTTGACTATCATGATCGAGCGTTAT 683  
Db 15958 ACCTGTTTTCATTCGCAATTTGTAATGGCGATTATGATGTTCCCGTTGAGGAAAGCGAACA 15899  
Qy 684 TGACGAGATCAAAATCGAAGTATCCAAACCTGCAACATGTTATTGACGCTGTGGAGCGA 743  
Db 15898 TATCATGAATCATGTTCCCACTTCAACATTTGAACCGGTGAATAAAGCAGCATGATTT 15839  
Qy 744 AGATAGTATCCCGAGCGCTATAAAGTCACAGCAGATAGTCTACCTGCCCATTTATAGA 803  
Db 15838 TGATCGTAGACCTAATGACGAGCAATCACTATTATCGTAAAGTCTGTTGTTCTTAAA 15779  
Qy 804 AGTGGTTTCAATGACCAATTCGAAAGCATTCCTGGAAGAAATCAGAAAAGATATGTTAAAA 862  
Db 15778 CGCTATAACAATGATGTAATATATATATATTTTAAAAAACTCAATTTATCATCATGATA 15720

## RESULT 7

US-10-029-386-25419/c

; Sequence 25419, Application US/10029386  
; Publication No. US20030194704A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Hanzel, David R.  
; APPLICANT: Hanzel, David R.  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR

```

; TITLE OF INVENTION:  EXPRESSION ANALYSIS TWO
; FILE REFERENCE:  AEOMICA-X-2
; CURRENT APPLICATION NUMBER:  US/10/029,386
; CURRENT FILING DATE:  2001-12-20
; NUMBER OF SEQ ID NOS:  34288
; SOFTWARE:  Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 25419
; LENGTH:  572
; TYPE:  DNA
; ORGANISM:  Homo sapiens
; FEATURE:
; OTHER INFORMATION:  MAP TO AL079336.5
; OTHER INFORMATION:  EXPRESSED IN HSLA, SIGNAL = 0.62
; OTHER INFORMATION:  NT HIT:  AF128626.1, EVALUATE 1.40e-01
US-10-029-386-25419

```

Query Match	3.7%;	Score 41;	DB 13;	Length 572;
Best Local Similarity	45.7%;	Pred. No. 2.1;		
Matches 143;	Conservative 0;	Mismatches 170;	Indels 0;	Gaps 0;

  

QY	7	GTTCACCACTCAAAAAGCGTCATCATTTGAAGGTGACAAAGCTGTGTTAAACAGAT	66
Db	429	GTTGCTGATGCTGATGTGGTGTGTCAGTATTGTTGATGTGGTGTGATGTTGGTGTT	370
QY	67	GTCTCAGTTCACGAATTAAGAGGGGTACAGCCTTGGTGAAGTTTGAGGCTGTGCTGGT	126
Db	369	GATGTAGTGGTGTGTGATATTGTTGATGTTTGATGTTGATGGTGGTGGTGT	310
QY	127	AACCCAACTGATTGGAAGCATATTGCTTTATAAGATTGGTCCAGAAGGTTCAATTCTAGGA	186
Db	309	GATGTAAGTGGTGTCAATATTGTTGATGTTAGTGGTGTGATGGTGGTATTGATGTT	250
QY	187	TGTGACATTGCTGGTACAGTGTCAAACTTGGACCAAAATGCTAGTACTGACTTGAAGGTT	246
Db	249	GCTGATGCTGATGTTGGTGGTGTCAATATTGTTGATGTTGGTGGTATGTTGATGTTCTGAT	190
QY	247	GGAGATACCGGTTTCGGTTTTGTTTCACGGTGTCTCCCAAAACAGATCCTAAAAATGGTGCA	306
Db	189	GCTGATGTTGGTGGTGTCAATGTTTGATGTTGATGGCGGTACTGATGTTCCCTGATGCTGAT	130
QY	307	TTTGCTGGAATG	319
Db	129	GATAGTGAATG	117

```

RESULT 8
US-10-312-841-1/c
; Sequence 1, Application US/10312841
; Publication No. US20030186277A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHC
; FILE REFERENCE: E01/1208/WO
; CURRENT APPLICATION NUMBER: US/10/312,841
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 2
; SEQ ID NO 1
; LENGTH: 3673778
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (3294164)
;
US-10-312-841-1

Query Match 3.7%; Score 41; DB 13; Length 3673778;
Best Local Similarity 47.5%; Pred. No. 3.2e+02;
Matches 122; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

QY 857 TTAAAAATTGATATTACTTTTGTATCGTGCATCTCGTCAAGAAATTTCTATTGGGTGCAA 916
|||||

```

Db	3131933	TTAAAAACAATACCGTATATATTTTTTCTTCTATATAAACACATATATCATTAATC	3131874
Qy	917	CAAGATTTTCGTAGTCAGCATATCATGAAGCCACAGTTAAATTCGTTAAGTTTATAA	976
Db	3131873	CATTTTATACTACTATAACAAAAATATCACAACCTAATAACAATAACAAAAATTTATTAA	3131814
Qy	977	ATCCACACCTTAACACGGTGATATCCATCATATGAATATTAAAGTTTTCAGCAACGGCT	1036
Db	3131813	CTTATATTTCTATAAACTAAAAAATCCAAATACTAAACATCTAAAAAAAACCTTT	3131754
Qy	1037	TAGATCATGTCCAGCTCTCACTGAAGGTATATAAGAGGTAAAAACAAAAATGTTAAAGT	1096
Db	3131753	TTTATTACATCATCACAAAAAAAATAAAAAAATAACAAAAAACAACAAAAAACTA	3131694
Qy	1097	ATGTTGCCAGGTTATAA 1113	
Db	3131693	AACTTACCCCTTTTATAA 3131677	

```

RESULT 9
US-10-312-841-2/c
; Sequence 2, Application US/10312841
; Publication No. US20030186277A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametere
; FILE REFERENCES: E01/1208/WO
; CURRENT APPLICATION NUMBER: US/10/312,841
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 2
; SEQ ID NO 2
; LENGTH: 3673778
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (379615)
US-10-312-841-2

```

	Query Match	3.6%;	Score 40.6;	DB 13;	Length 3673778;
	Best Local Similarity	49.3%;	Pred. No. 3.8e+02;		
	Matches 106;	Conservative 0;	Mismatches 109;	Indels 0;	Gaps 0;
Qy	896	AAGAAATCTTATGGGTGCAACAAGATTTCCTGCTAGTCCAGAAATATCATGAAGCCACAG	955		
Db	694104	AAAAATTAATTAATAAAAAATATATACAAAAACCAATATATATTTAAAAACAATTT	594045		
Qy	956	TTAAATTCGTTAAAGTTTATAAATCCACACCTTTAACACGGTGATATCCATCATATGAATA	1015		
Db	694044	AAAAATAACTTTAAAAAATAATATCTTCAAATAAAATAAATAATATACAATAATAAAAA	693985		
Qy	1016	TTAAAGTTTTTCAGCAACGGCTTAGATCATGTCCCGCTCTCTACTGAAGGTATATAAAGAG	1075		
Db	693984	ACAAAAATTTAAAAAACCAAAACCAATATATTTAAAAAACTCTTACAATATAAAAAAAAAT	693925		
Qy	1076	GTAAAAACAAAAATGTTAAAGTATGTTCGCCAGGTTA	1110		
Db	693924	AAAAATAAAATAACATAAAAAATTTATCAACAATTTA	693890		

RESULT 10  
US-09-864-761-20174/c  
; Sequence 20174, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY

FILE REFERENCE: Acomica-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIOR FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263.6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 20174  
LENGTH: 439  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AL035419.9  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.3  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.2  
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.4  
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.1  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.5  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.6  
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.5  
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.5  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.2  
US-09-864-761-20174  
Query Match 3.6%; Score 40.2; DB 9; Length 439;  
Best Local Similarity 46.8%; Pred. No. 3;  
Matches 126; Conservative 0; Mismatches 143; Indels 0; Gaps 0;  
QY 36 TGAAGGTGACAAAGCTGTTGTTAAACAGAGTCTCAGTTCCAGAAATTAAGGAGGTAC 95  
DB 385 TGATGGTGGTATGGTATGCTGCTGATGATGATGGTGGGATTAATGATGATGTC 326  
QY 96 AGCTTGTGTAAGGTTCAGGCTGTTGCTGGTAACCAACTGATTGGAAGCATATTGCTTA 155  
DB 325 TGCCGATGGTATGGTGGTATGATGATGGTATGCTGCTGATGATGATGATGATGATG 266  
QY 156 TAAGATTGGTCCAGAGTTCAATTTCTAGGATGTGACATTTGCTGGTACAGTTGCAAACT 215  
FILE REFERENCE: US-10-077-584-3/c  
SEQUENCE 3, Application US/10077584  
Publication No. US20030073610A1  
GENERAL INFORMATION:  
APPLICANT: LINDQUIST, SUSAN  
APPLICANT: KROBITSCH, SYLVIA  
APPLICANT: OUTEIRO, TIAGO P.  
TITLE OF INVENTION: YEAST SCREENS FOR THE TREATMENT OF HUMAN DISEASE  
FILE REFERENCE: ARCD:367US  
CURRENT APPLICATION NUMBER: US/10/077,584  
CURRENT FILING DATE: 2002-02-15  
PRIOR APPLICATION NUMBER: 60/269,157  
PRIOR FILING DATE: 2001-02-15  
NUMBER OF SEQ ID NOS: 9  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 3  
LENGTH: 1236  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)..(513)  
US-10-077-584-3  
Query Match 3.6%; Score 40; DB 15; Length 1236;  
Best Local Similarity 45.7%; Pred. No. 6.2;  
Matches 139; Conservative 0; Mismatches 165; Indels 0; Gaps 0;  
QY 25 GCCGTATCATTCGAAGTGACAAAGCTGTTGTTAAACAGATCTCTCAGTTCAGATTGA 84  
DB 346 GCTGTTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 287  
QY 85 AAGGAGGTACAGCCTTGGTGAAGGTTGAGGCTGTTGCTGTAACCAACTGATTGGAAG 144  
DB 286 GTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 227  
QY 145 CATATTGCTTATAAGATTGGTCCAGAGGTTCAATTCTAGGATGTGACATTGCTGTPACA 204  
DB 226 GTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 167  
QY 205 GTTGCTCAACTTGGACCAAAATGCTAGTACTGATGATGATGATGATGATGATGATGAT 264  
DB 166 GCTGTTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 107  
QY 265 TTGTTTCAAGGCTGCTTCCCAACAGATCCTAAAAAGTTCATTTGCTGAATATGCCAGG 324  
DB 106 GTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 47  
QY 325 GTTT 328  
DB 46 TTTT 43  
RESULT 12  
US-10-292-798-1825  
SEQUENCE 1825, Application US/10292798  
Publication No. US20030235833A1  
GENERAL INFORMATION:  
APPLICANT: SUWA, MAKIKO  
APPLICANT: ASAI, KIYOSHI  
APPLICANT: AKIYAMA, YUTAKA



```

; Publication No. US20030143668A1
;
; GENERAL INFORMATION:
;
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ASURATANI, HIROYUKI
;
; TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
;
; FILE REFERENCE: 084335/0152
;
; CURRENT APPLICATION NUMBER: US/10/017,161
;
; CURRENT FILING DATE: 2002-12-18
;
; PRIOR APPLICATION NUMBER: JP 2001/246789
;
; PRIOR FILING DATE: 2001-06-18
;
; NUMBER OF SEQ ID NOS: 2430
;
; SOFTWARE: PatentIn Ver. 2.1
;
; SEQ ID NO 2179
;
; LENGTH: 1168
;
; TYPE: DNA
;
; ORGANISM: Homo sapiens
;
; FEATURE:
;
; NAME/KEY: source
;
; LOCATION: (1)..(1168)
;
; FEATURE:
;
; NAME/KEY: CDS
;
; LOCATION: (201)..(968)
;
; US-10-017-161-2179

```

RESULT 16  
US-09-814-353-20029/c  
; Sequence 20029, Application US/09814353  
; Publication No. US20030165831A1  
; GENERAL INFORMATION:  
; APPLICANT: Lee, John  
; APPLICANT: Thompson, Pamela  
; APPLICANT: Lillie, James  
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR  
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND  
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER  
; FILE REFERENCE: MRI-006B  
; CURRENT APPLICATION NUMBER: US/09/814,353  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: US 60/191,031  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: US 60/207,124  
; PRIOR FILING DATE: 2000-05-25  
; PRIOR APPLICATION NUMBER: US 60/211,940  
; PRIOR FILING DATE: 2000-06-15  
; PRIOR APPLICATION NUMBER: US 60/216,820  
; PRIOR FILING DATE: 2000-07-07

```

; PRIOR APPLICATION NUMBER: US 60/220,561
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20029
; LENGTH: 10668
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 10556..10593, 10616, 10641, 10668
; OTHER INFORMATION: n = A,T,C or G
US-09-814-353-20029

```

```

RESULT 17
US-09-790-988-1/c
; Sequence 1, Application US/09790988
; Patent No. US20020127687A1
; GENERAL INFORMATION:
; APPLICANT: SHIGENOBU, SHUJI
; APPLICANT: WATANABE, HIDEMI
; APPLICANT: HATTORI, MASAHIRA
; APPLICANT: SAKAKI, YOSHIYUKI
; TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
; FILE REFERENCE: 081356/0159
; CURRENT APPLICATION NUMBER: US/09/790,988
; CURRENT FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: JP2000-107160
; PRIOR FILING DATE: 2000-04-07
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 640681
; TYPE: DNA
; ORGANISM: Buchnera sp.
US-09-790-988-1

```

	Query Match	3.4%	Score 37.6;	DB 10;	Length 640681;
	Best Local Similarity	48.6%;	Pred. No. 1e+03;		
	Matches 103;	Conservative 0;	Mismatches 109;	Indels 0;	Gaps 0;
Qy	716 AACATGTTATTGACGCTGTGGGAAGCGAAGATAGTATATCCCGAGGCCCTATAAAGTCACAG	775			
Db	547422 AATTAGTATTGAGATGCGAACTAACCGGAAGTATTGATCCAGAAGAAGCTATTTCGACGAG	5473			
Oy	776 CAGATAGCTACCTGCCACATTATTAGAAGTGGTTCCAATGACCATTTGAAGAGCATTCCTG	835			

Db 547362 CAGCTACTATTTAGCAGAACCACTAGAACATTTGTTGATTTAAAGAGATATTCGTGCAAC 547303  
Qy 836 AAGAATCAGAAAGATATGTTAAATGATATATCTTTGTTGATCGTCATCTCGTC 895  
Db 547302 CTGAATTCAGAGAGAAACCTGAATTTGAACCTATTTGTTGATCGTCAGTAGATGATT 547243  
Qy 896 AAGAAATTTCTATTGGGTGCAACAAGATTTCCT 927  
Db 547242 TAGAATTTACAGTTGCTTCTGCTAATGCTTT 547211

## RESULT 18

US-10-029-386-25433  
; Sequence 25433, Application US/10029386  
; Publication No. US20030194704A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David K.  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO  
; FILE REFERENCE: AEOMICA-X-2  
; CURRENT APPLICATION NUMBER: US/10/029.386  
; CURRENT FILING DATE: 2001-12-20  
; NUMBER OF SEQ ID NOS: 34288  
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 25433  
; LENGTH: 573  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AL133232.9  
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.16  
; OTHER INFORMATION: NT HIT: U63807.1, EVALUE 8.50e+00  
US-10-029-386-25433

Query Match 3.4%; Score 37.4; DB 13; Length 573;  
Best Local Similarity 45.4%; Pred. No. 19;  
Matches 134; Conservative 0; Mismatches 161; Indels 0; Gaps 0;

Qy 36 TGAAGGTGCAAAAGCTGTTGTTAAACACAGATGCTCAGTTCACAGATTAAGGAGGTAC 95  
Db 60 TGGTGGCGATATGCTGATGCAATGCTGGCGATGATGATGATAAGGTGATGCTGTC 119  
Qy 96 AGCTTGGTGAAGTGTGAGCTGTTGCTGTTAAACCACTGATTTGGAAGCATATGCTTA 155  
Db 120 TGCTGCTGTTAATGGTAGTGATGATGATGATGATGATGATGATGATGATGATGATG 179  
Qy 156 TAAGATTGGTCCAGAGGTTCAATCTAGGATGTGACATGCTGTCAGTTGTCAACT 215  
Db 180 TGATGATGGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 239  
Qy 216 TGGACCAAACTAGTACTGCTTCAAGGTTGGAGATACCGGTTTCGGTTTTCACGG 275  
Db 240 TGATGATCATGGTGTACAGATGCTGTTAATGATGATGATGATGATGATGATGATGATG 299  
Qy 276 TGCTTCCCAACACAGCTTAAATGATGATGATGATGATGATGATGATGATGATGATGAT 330  
Db 300 TGATGGTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 354

## RESULT 19

US-10-029-386-22786/c  
; Sequence 22786, Application US/10029386  
; Publication No. US20030194704A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David K.  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO  
; FILE REFERENCE: AEOMICA-X-2

; CURRENT APPLICATION NUMBER: US/10/029.386  
; CURRENT FILING DATE: 2001-12-20  
; NUMBER OF SEQ ID NOS: 34288  
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 22786  
; LENGTH: 835  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AC006549.28  
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.8  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.2  
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.2  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.3  
; OTHER INFORMATION: SWISSPROT HIT: O29522, EVALUE 1.80e+00  
US-10-029-386-22786

Query Match 3.4%; Score 37.4; DB 13; Length 835;  
Best Local Similarity 48.0%; Pred. No. 24;  
Matches 107; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

Qy 107 AGGTGAGGCTGTGCTGTGTAACCCAACTGATTTGGAAGCATATGCTTATAAGATTGGTC 166  
Db 568 AGCATGAGGATGCTGATCCTGGCTCGGTGATGATGATGATGATGATGATGATGATGATG 509  
Qy 167 CAGAAAGTTCAATCTAGGATGTGACATTCCTGCTGACATGTTGCAAACTTGGACCAATG 226  
Db 508 GTGATGGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 449  
Qy 227 CTAGTACTGACTTTGAAGTTGGAGATACCGGTTTCGGTTTTCGTTTTCACGGTCTTCCCAA 286  
Db 448 ATAGTGTGTAATGAGATGTAATGATGATGATGATGATGATGATGATGATGATGATGATG 389  
Qy 287 CAGATCTAAAAATGTTGCTGATTTGCTGAATATCCAGGGTTTA 329  
Db 388 CTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 346

## RESULT 20

US-10-369-493-42193  
; Sequence 42193, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369.493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 42193  
; LENGTH: 1200  
; TYPE: DNA  
; ORGANISM: Lactococcus lactis  
US-10-369-493-42193

Query Match 3.3%; Score 37.2; DB 12; Length 1200;  
Best Local Similarity 48.6%; Pred. No. 34;  
Matches 102; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

Qy 900 AATTCTATTGGGTGCAACAGATTTCTGCTAGTCCAGATATCATGACCCACAGTTAA 959  
Db 165 AATTACAAATTGGAACCAACAGGACTAGTCTGCTGAGAGAGATAAAAAAGTCAACTTTA 224  
Qy 960 ATTCTGTTAAGTTTATAAATCCACACCTTAACACCGGTGATATCCATCATATGAATATA 1019

Db 225 CTCGAAAAATTTTCCTAAATTTAGCGTCATTTGAATTTGATTTAGATGAAGTTGAAAAAGAA 284  
Qy 1020 AGTTTTCAGCAACGGCTTAGATGATGCCAGCTCTCACTGAAGGTATAAAAGAGTAA 1079  
Db 285 AGATGGTGAGCTTTGGTCTAATTTATGTTAAAGGATGATGTCATGCTTAAAGGAGCAGG 344  
Qy 1080 AACAACAAAATGTTAAAGTATGTTGCCAGTT 1109  
Db 345 ATATGAATTCACAAAGGTTTTCGAATGTT 374

## RESULT 21

US-10-252-157-159/c  
; Sequence 159, Application US/10252157  
; Publication No. US20030190640A1  
; GENERAL INFORMATION:  
; APPLICANT: Paris, Mary  
; APPLICANT: Pearson, Cecelia I.  
; TITLE OF INVENTION: GENES EXPRESSED IN PROSTATE CANCER  
; FILE REFERENCE: PA-0027-1 US  
; CURRENT APPLICATION NUMBER: US/10/252.157  
; CURRENT FILING DATE: 2002-10-01  
; PRIOR APPLICATION NUMBER: 60/295,048  
; PRIOR FILING DATE: 2001-05-31  
; NUMBER OF SEQ ID NOS: 501  
; SOFTWARE: PERL Program  
; SEQ ID NO 159  
; LENGTH: 7536  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID No. US20030190640A1 903479.3  
; LOCATION: 3017-3175, 5473, 5489, 5500, 5504, 6570  
; OTHER INFORMATION: a, t, c, g, or other  
US-10-252-157-159

Query Match 3.3%; Score 37.2; DB 13; Length 7536;  
Best Local Similarity 51.9%; Pred. No. 1e+02;  
Matches 84; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

Qy 936 AGAATATCATGAAGCCACAGTAAATTCGTTAAAGTTTATAATCCACACCTTAAACAACGG 995  
Db 2834 ATATTTCCCTGAATATGCAATTTAAAAAATAAAGTTTATGAGACAACATATGATTAAGT 2775  
Qy 996 TGATATCCATCATATGAATATAAAGTTTTCAGCAACGGCTTAGATGATGTCACAGCTCT 1055  
Db 2774 CAATTTCCACAGTTATAAATATTTCAATTTACAGCTACACACTATAAATTTACAGGAATAA 2715  
Qy 1056 CACTGAAGGTATAAAGCAAGGTAAACAAACAAATGTTAAAGTA 1097  
Db 2714 AATTCAGTAAACAGAAAGAGGAGGGGAAATAAAGCA 2673

## RESULT 22

US-10-198-846-14020/c  
; Sequence 14020, Application US/10198846  
; Publication No. US2003009974A1  
; GENERAL INFORMATION:  
; APPLICANT: Lillie, James  
; APPLICANT: Xu, Yongyao  
; APPLICANT: Wang, Youzhen  
; APPLICANT: Steinmann, Kathleen  
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS  
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND  
; TITLE OF INVENTION: THERAPY OF BREAST CANCER  
; FILE REFERENCE: MRI-049  
; CURRENT APPLICATION NUMBER: US/10/198.846  
; CURRENT FILING DATE: 2002-07-18  
; PRIOR APPLICATION NUMBER: 60/306,220  
; PRIOR FILING DATE: 2001-07-18

; NUMBER OF SEQ ID NOS: 14084  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 14020  
; LENGTH: 7772  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-198-846-14020

Query Match 3.3%; Score 37.2; DB 15; Length 7772;  
Best Local Similarity 51.9%; Pred. No. 1e+02;  
Matches 84; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

Qy 936 AGAATATCATGAAGCCACAGTAAATTCGTTAAAGTTTATAAATCCACACCTTAAACAACGG 995  
Db 3446 ATATTTCCCTGAATATGCAATTTAAAAAATAAAGTTTATGAGACAACATATGAATTAAGT 3387  
Qy 996 TGATATCCATCATATGAATATAAAGTTTTCAGCAACGGCTTAGATGATGTCACAGCTCT 1055  
Db 3386 CAATTTCCACAGTTATAAATATTTCAATTTACAGCTACACACTATAAATTTACAGGAATAA 3327  
Qy 1056 CACTGAAGGTATAAAGCAAGGTAAACAAACAAATGTTAAAGTA 1097  
Db 3326 AATTCAGTAAACAGAAAGAGGAGGGGAAATAAAGCA 3285

## RESULT 23

US-09-788-711A-3/c  
; Sequence 3, Application US/09788711A  
; Patent No. US20020058328A1  
; GENERAL INFORMATION:  
; APPLICANT: Tania Tamsin Testa  
; TITLE OF INVENTION: NOVEL COMPOUNDS  
; FILE REFERENCE: GP-30225  
; CURRENT APPLICATION NUMBER: US/09/788.711A  
; CURRENT FILING DATE: 2001-02-20  
; PRIOR APPLICATION NUMBER: 0004196.2  
; PRIOR FILING DATE: 2000-02-19  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 3  
; LENGTH: 8772  
; TYPE: DNA  
; ORGANISM: HOMO SAPIENS  
US-09-788-711A-3

Query Match 3.3%; Score 37; DB 9; Length 8772;  
Best Local Similarity 58.7%; Pred. No. 1.2e+02;  
Matches 64; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

Qy 11 CAACCACTCAAAAAGCCGTCATCATTTGAAGGTGACAAAGCTGTTGTTAAAAACAGATGCT 70  
Db 5456 CAGCTGCAGGAATAGCTGTCCAGTCTGTTGCTGCAATAGCTGTTAGCAGGACACGGGTTT 5397  
Qy 71 CAGTCCAGAAATTAAGGAGGGTACAGCCTTGCTGAAGTTGAGGCTGT 119  
Db 5396 GAGTCACAGGGTCAGGAGGCTACAGCCTTGCTCCACGTTGATGCTCT 5348

## RESULT 24

US-09-788-711A-1/c  
; Sequence 1, Application US/09788711A  
; Patent No. US20020058328A1  
; GENERAL INFORMATION:  
; APPLICANT: Tania Tamsin Testa  
; TITLE OF INVENTION: NOVEL COMPOUNDS  
; FILE REFERENCE: GP-30225  
; CURRENT APPLICATION NUMBER: US/09/788.711A  
; CURRENT FILING DATE: 2001-02-20  
; PRIOR APPLICATION NUMBER: 0004196.2  
; PRIOR FILING DATE: 2000-02-19  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 1

```
; LENGTH: 8871
; TYPE: DNA
; ORGANISM: HOMO SAPIENS
US-09-788-711A-1

Query Match      3.3%; Score 37; DB 9; Length 8871;
Best Local Similarity 58.7%; Pred. No. 1.2e+02;
Matches 64; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

Qy 11 CAACCACTCAAAAGCGTCATCATTTGAAGTGCACAAAGCTGTTGTTAAACACAGATGCT 70
Db 5456 CAGCTGCAGGAATAGCTGTCCAGTCGTTGCTGCAATAGCTGTTAGCAGGACACGGGTTT 5397

Qy 71 CAGTTCCAGAAATTAAGGAGGAGGTACAGCCTTGTGTGAAGTTGAGGCTGT 119
Db 5396 GAGTCACAGAGGTACAGGAGGCTACAGCCTTGTCTCCAGCTTGATGCTCT 5348

RESULT 25
US-10-225-567A-523/c
; Sequence 523, Application US/10225567A
; Publication No. US20030113798A1
; GENERAL INFORMATION:
; APPLICANT: LifeSpan Biosciences
; APPLICANT: Brown, Joseph P.
; APPLICANT: Burmer, Glenna C.
; APPLICANT: Roush, Christine L.
; TITLE OF INVENTION: ANTIGENIC PEPTIDES AND ANTIBODIES FOR G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 1920-4-4
; CURRENT APPLICATION NUMBER: US/10/225.567A
; CURRENT FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: 60/257,144
; PRIOR FILING DATE: 2000-12-19
; NUMBER OF SEQ ID NOS: 232
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 523
; LENGTH: 10531
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-225-567A-523

Query Match      3.3%; Score 37; DB 15; Length 10531;
Best Local Similarity 58.7%; Pred. No. 1.4e+02;
Matches 64; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

Qy 11 CAACCACTCAAAAGCGTCATCATTTGAAGTGCACAAAGCTGTTGTTAAACACAGATGCT 70
Db 5518 CAGCTGCAGGAATAGCTGTCCAGTCGTTGCTGCAATAGCTGTTAGCAGGACACGGGTTT 5459

Qy 71 CAGTTCCAGAAATTAAGGAGGAGGTACAGCCTTGTGTGAAGTTGAGGCTGT 119
Db 5458 GAGTCACAGAGGTACAGGAGGCTACAGCCTTGTCTCCAGCTTGATGCTCT 5410

RESULT 26
US-10-292-798-931/c
; Sequence 931, Application US/10292798
; Publication No. US2003023583A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
; FILE REFERENCE: 084335/166
; CURRENT APPLICATION NUMBER: US/10/292.798
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: 10/017,161
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: JP 2001-246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2070
; SOFTWARE: PatentIn Ver. 2.1
```



LOCATION: (19815)..(19941)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (20054)..(20254)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (20547)..(20721)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (21048)..(21160)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (21338)..(21417)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (21507)..(21633)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (21721)..(21843)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (22399)..(22526)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (22761)..(22849)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (22954)..(23148)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (23287)..(23457)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (23557)..(23791)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (24143)..(24170)  
US-10-292-798-931

Query Match 3.3%; Score 37; DB 12; Length 24370;  
Best Local Similarity 58.7%; Pred. No. 2.3e+02;  
Matches 64; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 11 CAACCACTCAAAAGCGCTCATCTTGAAGGTCACAAAGCTGTGTGTTAAACAGAGTGCT 70  
DB 14741 CAGTGCAGGAATAGCTGTCCAGTCTGTGCTGCATAGCTGTACACGACACGGGTTT 14682  
QY 71 CAGTTCAGAAATTAAGGAGGGGTACAGCCTTGTGTGAAGTTGAGGCTGT 119  
DB 14681 GAGTCACAAGGGTCAGGAGGGGTACAGCCTTGTCTCCACGCTTGATGCTCT 14633

RESULT 27  
US-09-880-107-3814/c  
Sequence 3814, Application US/09880107  
Patent No. US20020142981A1  
GENERAL INFORMATION:  
APPLICANT: Horne, Darci T.  
APPLICANT: Vockley, Joseph G.  
APPLICANT: Scherf, Uwe  
APPLICANT: Gene Logic, Inc.  
TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer  
FILE REFERENCE: 44921-5028-WO  
CURRENT APPLICATION NUMBER: US/09/880,107  
PRIOR FILING DATE: 2001-06-14  
PRIOR APPLICATION NUMBER: US 60/211,379  
PRIOR FILING DATE: 2000-06-14  
PRIOR APPLICATION NUMBER: US 60/237,054  
PRIOR FILING DATE: 2000-10-02  
NUMBER OF SEQ ID NOS: 3950  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 3814  
LENGTH: 198285

TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: Genbank Accession No. US20020142981A1 X87344  
US-09-880-107-3814

Query Match 3.3%; Score 37; DB 10; Length 198285;  
Best Local Similarity 49.2%; Pred. No. 7.7e+02;  
Matches 97; Conservative 0; Mismatches 100; Indels 0; Gaps 0;  
QY 829 ATTCTCGAAGAATCAGAAAGATAATGTTAAATGATATTACTTTGTTGTTGTCGCA 888  
DB 169231 ATTCATTTTAAAGAAATATATAATAATGTAATACTTTTCCCACTCCACTTGC 169172  
QY 889 TCTGGTCAAGAAATCTATTGGTGCAACAAGATTTCTGCTAGTCCAGAATATCATGAA 948  
DB 169171 TTTTCTAAGCAGTATTATGATTTTCACTTTTACAGTTTAAATACATATATTAG 169112  
QY 949 GCCACAGTTAAATTCGTTAAGTTTATAATCCACACCTTTTAAACCGGTGATATCCATCAT 1008  
DB 169111 GTGTAGGTTTCTTGTGTTTATTAATACAGGATTACCTACAGAGGAACATAACAT 169052  
QY 1009 ATGAATATTAAGTTT 1025  
DB 169051 AGTAATATAATAAATT 169035

RESULT 28  
US-09-864-761-3403/c  
Sequence 3403, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE REFERENCE: Aemica-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263.6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21

RESULT 29  
 US-09-938-842A-243/c  
 ; Sequence 243, Application US/09938842A  
 ; Patent No. US20020160378A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Harper, Jeff  
 ; APPLICANT: Kreps, Joel  
 ; APPLICANT: Wang, Xun  
 ; APPLICANT: Zhu, Tong  
 ; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING  
 ; TITLE OF INVENTION: SAME, AND METHODS OF USE  
 ; FILE REFERENCE: SCRIPI300-3  
 ; CURRENT APPLICATION NUMBER: US/09/938,842A  
 ; CURRENT FILING DATE: 2001-08-24  
 ; PRIOR APPLICATION NUMBER: US 60/227,866  
 ; PRIOR FILING DATE: 2000-08-24  
 ; PRIOR APPLICATION NUMBER: US 60/264,647  
 ; PRIOR FILING DATE: 2001-01-16  
 ; PRIOR APPLICATION NUMBER: US 60/300,111  
 ; PRIOR FILING DATE: 2001-06-22  
 ; NUMBER OF SEQ ID NOS: 5379  
 ; SEQ ID NO 243  
 ; LENGTH: 3150  
 ; TYPE: DNA  
 ; ORGANISM: Arabidopsis thaliana  
 ; US-09-938-842A-243  
 Query Match 3.3%; Score 36.8; DB 10; Length 3150;

US-09-815-242-6606

Query Match	3.3%	Score 36.6;	DB 9;	Length 969;
Best Local Similarity	58.9%	Pred. No. 43;		
Matches 63;	Conservative 0;	Mismatches 44;	Indels 0;	Gaps 0;

  

Qy	37	GAAGGTGACAAAGCTGTTGTTAAACACAGATGCTCTCAGTTCAGATTCCAGAAATTAAAGGAGGTACA	96
Db	55	GGAAAGTAAAGAGGTACTTTGAAGAAGCAGAGTAACACTGCCAGAAATTATCGGAACATCAA	114
Qy	97	GCCTTGGTGAAGGTTGAGGCTGTTGCTGGTAACCCCACTGATTGGAA	143
Db	115	GTGTTGGTAAAGAATACCGACATCAATTAATCCGATTGATTGGAA	161

Search completed: January 10, 2004, 10:03:48  
Job time : 477 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 10, 2004, 06:18:49 ; Search time 2773 Seconds

(without alignments)  
9755.096 Million cell updates/sec

Title: US-10-081-644-1

Perfect score: 1113

Sequence: 1 atgcagttccaaaccactca.....agatgttcagggtataa 1113

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 90 summaries

Database :

EST:

1: em\_estba:\*\*

2: em\_esthum:\*\*

3: em\_estin:\*\*

4: em\_estmu:\*\*

5: em\_estov:\*\*

6: em\_estopl:\*\*

7: em\_estro:\*\*

8: em\_estc:\*\*

9: gb\_est1:\*\*

10: gb\_est2:\*\*

11: gb\_estc:\*\*

12: gb\_estd:\*\*

13: gb\_este:\*\*

14: gb\_estf:\*\*

15: em\_estfun:\*\*

16: em\_estom:\*\*

17: em\_gss\_hum:\*\*

18: em\_gss\_inv:\*\*

19: em\_gss\_pln:\*\*

20: em\_gss\_vrt:\*\*

21: em\_gss\_fun:\*\*

22: em\_gss\_man:\*\*

23: em\_gss\_mus:\*\*

24: em\_gss\_pro:\*\*

25: em\_gss\_rod:\*\*

26: em\_gss\_phg:\*\*

27: em\_gss\_vrl:\*\*

28: gb\_gss1:\*\*

29: gb\_gss2:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	294.4	26.5	1011	29	CNS06N13	AL406189 T7 end of
C 2	182.4	16.4	896	29	CNS06LDW	AL404058 T3 end of
C 3	181.2	16.3	881	29	CNS06LDV	AL404057 T7 end of
C 4	154.6	13.9	1072	29	CNS06LTC	AL404614 T7 end of
-----						
C 1	294.4	26.5	1011	29	CNS06N13	AL406189 T7 end of
C 2	182.4	16.4	896	29	CNS06LDW	AL404058 T3 end of
C 3	181.2	16.3	881	29	CNS06LDV	AL404057 T7 end of
C 4	154.6	13.9	1072	29	CNS06LTC	AL404614 T7 end of
-----						
C 1	294.4	26.5	1011	29	CNS06N13	AL406189 T7 end of
C 2	182.4	16.4	896	29	CNS06LDW	AL404058 T3 end of
C 3	181.2	16.3	881	29	CNS06LDV	AL404057 T7 end of
C 4	154.6	13.9	1072	29	CNS06LTC	AL404614 T7 end of
-----						
C 1	294.4	26.5	1011	29	CNS06N13	AL406189 T7 end of
C 2	182.4	16.4	896	29	CNS06LDW	AL404058 T3 end of
C 3	181.2	16.3	881	29	CNS06LDV	AL404057 T7 end of
C 4	154.6	13.9	1072	29	CNS06LTC	AL404614 T7 end of
-----						
C 1	294.4	26.5	1011	29	CNS06N13	AL406189 T7 end of
C 2	182.4	16.4	896	29	CNS06LDW	AL404058 T3 end of
C 3	181.2	16.3	881	29	CNS06LDV	AL404057 T7 end of
C 4	154.6	13.9	1072	29	CNS06LTC	AL404614 T7 end of
-----						
C 1	294.4	26.5	1011	29	CNS06N13	AL406189 T7 end of
C 2	182.4	16.4	896	29	CNS06LDW	AL404058 T3 end of
C 3	181.2	16.3	881	29	CNS06LDV	AL404057 T7 end of
C 4	154.6	13.9	1072	29	CNS06LTC	AL404614 T7 end of
-----						
C 1	294.4	26.5	1011	29	CNS06N13	AL406189 T7 end of
C 2	182.4	16.4	896	29	CNS06LDW	AL404058 T3 end of
C 3	181.2	16.3	881	29	CNS06LDV	AL404057 T7 end of
C 4	154.6	13.9	1072	29	CNS06LTC	AL404614 T7 end of
-----						
C 1	294.4	26.5	1011	29	CNS06N13	AL406189 T7 end of
C 2	182.4	16.4	896	29	CNS06LDW	AL404058 T3 end of
C 3	181.2	16.3	881	29	CNS06LDV	AL404057 T7 end of
C 4	154.6	13.9	1072	29	CNS06LTC	AL404614 T7 end of
-----						
C 1	294.4	26.5	1011	29	CNS06N13	AL406189 T7 end of
C 2	182.4	16.4	896	29	CNS06LDW	AL404058 T3 end of
C 3	181.2	16.3	881	29	CNS06LDV	AL404057 T7 end of
C 4	154.6	13.9	1072	29	CNS06LTC	AL404614 T7 end of
-----						
C 1	294.4	26.5	1011	29	CNS06N13	AL406189 T7 end of
C 2	182.4	16.4	896	29	CNS06LDW	AL404058 T3 end of
C 3	181.2	16.3	881	29	CNS06LDV	AL404057 T7 end of
C 4	154.6	13.9	1072	29	CNS06LTC	AL404614 T7 end of
-----						
C 1	294.4	26.5	1011	29	CNS06N13	AL406189 T7 end of
C 2	182.4	16.4	896	29	CNS06LDW	AL404058 T3 end of
C 3	181.2	16.3	881	29	CNS06LDV	AL404057 T7 end of
C 4	154.6	13.9	1072	29	CNS06LTC	AL404614 T7 end of
-----						
C 1	294.4	26.5	1011	29	CNS06N13	AL406189 T7 end of
C 2	182.4	16.4	896	29	CNS06LDW	AL404058 T3 end of
C 3	181.2	16.3	881	29	CNS06LDV	AL404057 T7 end of
C 4	154.6	13.9	1072	29	CNS06LTC	AL404614 T7 end of
-----						
C 1	294.4	26.5	1011	29	CNS06N13	AL406189 T7 end of
C 2	182.4	16.4	896	29	CNS06LDW	AL404058 T3 end of
C 3	181.2	16.3	881	29	CNS06LDV	AL404057 T7 end of
C 4	154.6	13.9	1072	29	CNS06LTC	AL404614 T7 end of
-----						
C 1	294.4	26.5	1011	29	CNS06N13	AL406189 T7 end of
C 2	182.4	16.4	896	29	CNS06LDW	AL404058 T3 end of
C 3	181.2	16.3	881	29	CNS06LDV	AL404057 T7 end of
C 4	154.6	13.9	1072	29	CNS06LTC	AL404614 T7 end of
-----						
C 1	294.4	26.5	1011	29	CNS06N13	AL406189 T7 end of
C 2	182.4	16.4	896	29	CNS06LDW	AL404058 T3 end of
C 3	181.2	16.3	881	29	CNS06LDV	AL404057 T7 end of
C 4	154.6	13.9	1072	29	CNS06LTC	AL404614 T7 end of
-----						
C 1	294.4	26.5	1011	29	CNS06N13	AL406189 T7 end of
C 2	182.4	16.4	896	29	CNS06LDW	AL404058 T3 end of
C 3	181.2	16.3	881	29	CNS06LDV	AL404057 T7 end of
C 4	154.6	13.9	1072	29	CNS06LTC	AL404614 T7 end of
-----						
C 1	294.4	26.5	1011	29	CNS06N13	AL406189 T7 end of
C 2	182.4	16.4	896	29	CNS06LDW	AL404058 T3 end of
C 3	181.2	16.3	881	29	CNS06LDV	AL404057 T7 end of
C 4	154.6	13.9	1072	29	CNS06LTC	AL404614 T7 end of
-----						
C 1	294.4	26.5	1011	29	CNS06N13	AL406189 T7 end of
C 2	182.4	16.4	896	29	CNS06LDW	AL404058 T3 end of
C 3	181.2	16.3	881	29	CNS06LDV	AL404057 T7 end of
C 4	154.6	13.9	1072	29	CNS06LTC	AL404614 T7 end of
-----						
C 1	294.4	26.5	1011	29	CNS06N13	AL406189 T7 end of
C 2	182.4	16.4	896	29	CNS06LDW	AL404058 T3 end of
C 3	181.2	16.3	881	29	CNS06LDV	AL404057 T7 end of
C 4	154.6	13.9	1072	29	CNS06LTC	AL404614 T7 end of
-----						
C 1	294.4	26.5	1011	29	CNS06N13	AL406189 T7 end of
C 2	182.4	16.4	896	29	CNS06LDW	AL404058 T3 end of
C 3	181.2	16.3	881	29	CNS06LDV	AL404057 T7 end of
C 4	154.6	13.9	1072	29	CNS06LTC	AL404614 T7 end of
-----						
C 1	294.4	26.5	1011	29	CNS06N13	AL406189 T7 end of
C 2	182.4	16.4	896	29	CNS06LDW	AL404058 T3 end of
C 3	181.2	16.3	881	29	CNS06LDV	AL404057 T7 end of
C 4	154.6	13.9	1072	29	CNS06LTC	AL404614 T7 end of
-----						
C 1	294.4	26.5	1011	29	CNS06N13	AL406189 T7 end of
C 2	182.4	16.4	896	29	CNS06LDW	AL404058 T3 end of
C 3	181.2	16.3	881	29	CNS06LDV	AL404057 T7 end of
C 4	154.6	13.9	1072	29	CNS06LTC	AL404614 T7 end of
-----						
C 1	294.4	26.5	1011	29	CNS06N13	AL406189 T7 end of
C 2	182.4	16.4	896	29	CNS06LDW	AL404058 T3 end of
C 3	181.2	16.3	881	29	CNS06LDV	AL404057 T7 end of
C 4	154.6	13.9	1072	29	CNS06LTC	AL404614 T7 end of
-----						
C 1	294.4	26.5	1011	29	CNS06N13	AL406189 T7 end of
C 2	182.4	16.4	896	29	CNS06LDW	AL404058 T3 end of
C 3	181.2	16.3	881	29	CNS06LDV	AL404057 T7 end of
C 4	154.6	13.9	1072	29	CNS06LTC	AL404614 T7 end of
-----						
C 1	294.4	26.5	1011	29	CNS06N13	AL406189 T7 end of
C 2	182.4	16.4	896	29	CNS06LDW	AL404058 T3 end of
C 3	181.2	16.3	881	29	CNS06LDV	AL404057 T7 end of
C 4	154.6	13.9	1072	29	CNS06LTC	AL404614 T7 end of
-----						
C 1	294.4	26.5	1011	29	CNS06N13	AL406189 T7 end of
C 2	182.4	16.4	896	29	CNS06LDW	AL404058 T3 end of
C 3	181.2	16.3	881	29	CNS06LDV	AL404057 T7 end of
C 4	154.6	13.9	1072	29	CNS06LTC	AL404614 T7 end of
-----						
C 1	294.4	26.5	1011	29	CNS06N13	AL406189 T7 end of
C 2	182.4	16.4	896	29	CNS06LDW	AL404058 T3 end of
C 3	181.2	16.3	881	29	CNS06LDV	AL404057 T7 end of
C 4	154.6	13.9	1072	29	CNS06LTC	AL404614 T7 end of
-----						
C 1	294.4	26.5	1011	29	CNS06N13	AL406189 T7 end of
C 2	182.4	16.4	896	29	CNS06LDW	AL404058 T3 end of
C 3	181.2	16.3	881	29	CNS06LDV	AL404057 T7 end of
C 4	154.6	13.9	1072	29	CNS06LTC	AL404614 T7 end of
-----						
C 1	294.4	26.5	1011	29	CNS06N13	AL406189 T7 end of
C 2	182.4	16.4	896	29	CNS06LDW	AL404058 T3 end of
C 3	181.2	16.3	881	29	CNS06LDV	AL404057 T7 end of
C 4	154.6	13.9	1072	29	CNS06LTC	AL404614 T7 end of
-----						
C 1	294.4	26.5	1011	29	CNS06N13	AL406189 T7 end of
C 2	182.4	16.4	896	29	CNS06LDW	AL404058 T3 end of
C 3	181.2	16.3	881	29	CNS06LDV	AL404057 T7 end of
C 4	154.6	13.9	1072	29	CNS06LTC	AL404614 T7 end of
-----						
C 1	294.4	26.5	1011	29	CNS06N13	AL406189 T7 end of
C 2	182.4	16.4	896	29	CNS06LDW	AL404058 T3 end of
C 3	181.2	16.3	881	29	CNS06LDV	AL404057 T7 end of
C 4	154.6	13.9	1072	29	CNS06LTC	AL404614 T7 end of
-----						
C 1	294.4	26.5	1011	29	CNS06N13	AL406189 T7 end of
C 2	182.4	16.4	896	29	CNS06LDW	AL404058 T3 end of
C 3	181.2	16.3	881	29	CNS06LDV	AL404057 T7 end of
C 4	154.6	13.9	1072	29	CNS06LTC	AL404614 T7 end of
-----						
C 1	294.4	26.5	1011	29	CNS06N13	AL406189 T7 end of
C 2	182.4	16.4	896	29	CNS06LDW	AL404058 T3 end of
C 3	181.2	16.3	881	29	CNS06LDV	AL404057 T7 end of
C 4	154.6	13.9	1072	29	CNS06LTC	AL404614 T7 end of
-----						
C 1	294.4	26.5	1011	29	CNS06N13	AL406189 T7 end of
C 2	182.4	16.4	896	29	CNS06LDW	AL404058 T3 end of
C 3	181.2	16.3	881	29	CNS06LDV	AL404057 T7 end of
C 4	154.6	13.9	1072	29	CNS06LTC	AL404614 T7 end of
-----						
C 1	294.4	26.5	1011	29	CNS06N13	AL406189 T7 end of
C 2	182.4	16.4	896	29	CNS06LDW	AL404058 T3 end of
C 3	181.2	16.3	881	29	CNS06LDV	AL404057 T7 end of
C 4	154.6	13.9	1072	29	CNS06LTC	AL404614 T7 end of
-----						
C 1	294.4	26.5	1011	29	CNS06N13	AL406189 T7 end of
C 2	182.4	16.4	896	29	CNS06LDW	AL404058 T3 end of
C 3	181.2	16.3	881	29	CNS06LDV	AL404057 T7 end of
C 4	154.6	13.9	1072	29	CNS06LTC	AL404614 T7 end of
-----						
C 1	294.4	26.5	1011	29	CNS06N13	AL406189 T7 end of
C 2	182.4	16.4	896	29	CNS06LDW	AL404058 T3 end of
C 3	181.2	16.3	881	29	CNS06LDV	AL404057 T7 end of
C 4	154.6	13.9	1072	29	CNS06LTC	AL404614 T7 end of
-----						
C 1	294.4	26.5	1011	29	CNS06N13	AL406189 T7 end of
C 2	182.4	16.4	896	29	CNS06LDW	AL404058 T3 end of
C 3	181.2	16.3	881	29	CNS06LDV	AL404057 T7 end of
C 4	154.6	13.9	1072	29	CNS06LTC	AL404614 T7 end of
-----						
C 1	294.4	26.5	1011	29	CNS06N13	AL406189 T7 end of
C 2	182.4	16.4	896	29	CNS06LDW	AL404058 T3 end of
C 3	181.2	16.3	881	29	CNS06LDV	AL404057 T7 end of
C 4	154.6	13.9	1072	29	CNS06LTC	AL404614 T7 end of
-----						
C 1	294.4	26.5	1011	29	CNS06N13	AL406189 T7 end of
C 2	182.4	16.4	896	29	CNS06LDW	AL404058 T3 end of
C 3	181.2	16.3	881	29	CNS06LDV	AL404057 T7 end of
C 4	154.6	13.9	1072	29	CNS06LTC	AL404614 T7 end of
-----						
C 1	294.4	26.5	1011	29	CNS06N13	AL406189 T7 end of
C 2	182.4	16.4	896	29	CNS06LDW	AL404058 T3 end of
C 3	181.2	16.3	881	29	CNS06LDV	AL404057 T7 end of
C 4	154.6	13.9	1072	29	CNS06LTC	AL404614 T7 end of
-----						
C 1	294.4	26.5	1011			

```

78 42.4 3.8 1048 13 BX354430
c 79 42.2 3.8 994 13 BX414650
c 80 42 3.8 556 29 BZ869915
c 81 42 3.8 1101 29 CNS00001
c 82 41.8 3.8 326 14 C92181
c 83 41.8 3.8 1201 13 BX384819
c 84 41.6 3.7 308 9 A1374249
c 85 41.6 3.7 465 29 BZ955772
c 86 41.6 3.7 617 12 BJ371293
c 87 41.6 3.7 712 13 BQ914346
c 88 41.4 3.7 238 29 BZ927574
c 89 41.4 3.7 425 29 BZ913257
c 90 41.4 3.7 468 29 BZ914511

```

## ALIGNMENTS

```

RESULT 1
CNS06N13/c 1011 bp DNA linear GSS 17-JUN-2001
LOCUS T7 end of clone AU00A008F10 of library AU00A from strain CBS 3082
DEFINITION of Saccharomyces kluyveri, genomic survey sequence.
ACCESSION AL406189
VERSION AL406189.1 GI:12169887
KEYWORDS GSS.
SOURCE Saccharomyces kluyveri
ORGANISM Saccharomyces kluyveri

```

```

REFERENCE
AUTHORS Souciet J.L., Aigle M., Artiguenave F., Blandin G.,
Bolotin-Fukuhara M., Bon E., Brottier P., Casaregola S.,
de Montigny J., Dujon B., Durrens P., Lepingle A., Florents B.,
Malpertuy A., Neuveglise C., Ozier-Kalogeropoulos O., Potier S.,
Saurin W., Tekala F., Toffano-Nioche C., Wesolowski-Louvel M.,
Wincker P. and Weissenbach J.

```

```

TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of
yeast species for molecular evolution studies
JOURNAL FEBS Lett. 487 (1), 3-12 (2000)
MEDLINE 1152876
PUBMED 20584711
REFERENCE 2 (bases 1 to 1011)
AUTHORS Neuveglise C., Bon E., Lepingle A., Wincker P., Artiguenave F.,
Gaillardin C., and Casaregola S.

```

```

TITLE Genomic exploration of the hemiascomycetous yeasts: 9.
JOURNAL Saccharomyces kluyveri
MEDLINE FEBS Lett. 487 (1), 56-60 (2000)
PUBMED 20584719
REFERENCE 3 (bases 1 to 1011)
AUTHORS Genoscope.

```

```

TITLE Direct Submission
JOURNAL Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. [E-mail :
segref@genoscope.cns.fr - Web : www.genoscope.cns.fr]
COMMENT This GSS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.

```

```

FEATURES
source 1..1011
/organism="Saccharomyces kluyveri"
/mol_type="genomic DNA"
/strain="CBS 3082"
/db_xref="taxon:4934"
/clone="AU00A008F10"

```

```

misc_feature 271 a 222 c 215 g 295 t 8 others
/clone lib="AU00A"
/notes="end : T7"
complement (<185..>1007)
/notes="similar to Saccharomyces cerevisiae ORF YNL134c [
similarity to C.carbonum tox d gene ]
1 putative frameshift(s)"
/evidence=not experimental
misc_feature complement (<188..>1007)
/notes="similar to Saccharomyces cerevisiae ORF YCR102c [
similarity to C.carbonum tox d gene ]
1 putative frameshift(s)
similar to Saccharomyces cerevisiae ORF YLR460c [
similarity to C.carbonum tox d protein ]"
/evidence=not experimental

```

```

BASE COUNT 271 a 222 c 215 g 295 t
ORIGIN
Query Match 26.5%; Score 294.4; DB 29; Length 1011;
Best Local Similarity 61.3%; Pred. No. 2.1e-62;
Matches 504; Conservative 2; Mismatches 314; Indels 2; Gaps 2;

```

```

Qy 292 CCTAAATAATGGTCATTTGCTGAATATGCCAGGGTTTATCCACCTTTGTTTACAGAGT 351
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1004 CCTGAGAAATGGTGGCTTTGCTGAATACAGCCGATTGGTGGATTCCTTCTTATAAGGCA 945
Qy 352 AACTTAACCTCACTCAACTGCTGATGAATTTCTGAAGGCCCTGTGAAGACTTCGATCT 411
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
944 GAGTTAACACCTAGTACMAAGRAT-ACATACCAAGAAAGGTGTGGTCAAAACATTTGAGTCG 886
Qy 412 GCTGCATCATTTGCCAGTTCCTGTTGACAACTGCTGGTGTGTAGTTGTTCATCATCTGGGC 471
Db : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
885 GACTB-TCGCTACTCTGTCCTTGACCACGGCTGGTGGTCTTAAGTTCATCATCTGGGC 827
Qy 472 TCAAAATGGAATGGACCCATCTACCCGCAACATCACTCATCCATATATGATTGGGGT 531
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
826 AGCAAGTTTGAATGGGAACCATCAAAAGCCTCAACATGATTTCCCTCTCTGATTGGGGT 767
Qy 532 GGTGCTACAGAGTGGTCAACAACTAATCCAAAGTTGCCAACAATCATCGTTATCTACT 591
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
766 GGTGCTACAGAGTGGTCAATTTTGTGTTGCTAGTTGGCTAAGAAGATGACTGTGTACACA 707
Qy 592 AAGATTGTAACCTGCTCTTCTAAAGAGCATGAAAGCTTTTAAAGTCTTATGGTGTGAT 651
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
706 AAGATTGTTGGCGGTGCTTCCAAAGAAACAGRATCCATCTTGAAGCAATACGTTGCTGAT 647
Qy 652 GATGCTTTGACATCATGATGATGAGCGGTTATTGACAGATCAATATCGAAGTATCCAAAC 711
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
646 GAGATCTTCGATTACCATGATGAGGACGTTATTGAAACAAATAAGAGCCAAATACCCAAAT 597
Qy 712 CTGCAACATGTTATTGACGCTGTGGGAAGGAAGATAGTATCCCGAGGCCCTATTAAGTC 771
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
586 CTACAACATGTTATCGATGCTGTTTCTAACGTGCAAACTATCCACAGGTCTACAGATGT 527
Qy 772 ACAGCAGATAGTCTACTCTGCCACATTTATTAGAAGTGGTTCCCAATGACCATTTGAAAGCAT 831
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
526 ACTCCAGATAAATAGCAACCACTTGTGCAATTTGGCTGCTTTAAGAGGGGGACATTT 467
Qy 832 CTGGAAGAAATCAGAAAAGATATGTTAAATGATATTAATCTTTGTTGTTGATCGTCACT 891
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
466 GAAAAGAGAAGAAAGAAATCAAAATGTAAGATGGAGCAGCCCTGTTTATATCTTGGCTGCC 407
Qy 892 GGTCAAGAAATCTATTGGTGCACAAGATTTCTGCTAGTCCAGAAATATCATGAAGCC 951
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
406 GGTCAAGAGTGGCAATTCGGGCCCCCACTATCCCTGCTAAACCAAGAAATACAGAGGGCG 347
Qy 952 ACAGTTAAATTCGTTAAGTTTAAATCCACACCTTTAAACACCGGTGATATCCATCATATG 1011
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
346 ACTAGAGATTCGTGGAATTTGTCAACCCCAAGRATCAACACGTTGAACTTCACATCT 287
Qy 1012 AATATTAAAGTTTTCAGAACCGCTTAGATGATGTCCTCCAGCTCTCTCAAGAGGTATAAA 1071
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
286 CTAAATCCGTTTGTGAAAACCGGTTGGAAGATGTTCTCTGGACTTACTGTCGAATCCAA 227

```

```

Qy 1072 GAAGGTAACAAACAAAATGTTAAGTATGTTCCAGGTTATAA 1113
Db 226 GCAGGTAAGATTCGGGTGTCAAAATTTGTTGCAGCTTGAA 185

RESULT 2
CNS06LDW/c
LOCUS
DEFINITION
T3 end of clone AT0AA012G10 of library AT0AA from strain CBS 4311
of Saccharomyces servazzii, genomic survey sequence.
ACCESSION
AL404058
VERSION
AL404058.1 GI:12165095
KEYWORDS
GSS.
SOURCE
Saccharomyces servazzii
ORGANISM
Saccharomyces servazzii
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
REFERENCE
1 (bases 1 to 896)
Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Lorente,B.,
Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
Saurin,W., Tekaja,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
Wincker,P. and Weissenbach,J.
Genomic exploration of the hemiascomycetous yeasts: 1. A set of
yeast species for molecular evolution studies
FEBS Lett. 487 (1), 3-12 (2000)
20584711
11152876
2 (bases 1 to 896)
Casaregola,S., Lepingle,A., Bon,E., Neuveglise,C., Nguyen,H.,
Artiguenave,F., Wincker,P. and Gaillardin,C.
Genomic exploration of the hemiascomycetous yeasts: 7.
Saccharomyces servazzii
FEBS Lett. 487 (1), 47-51 (2000)
20584717
11152882
3 (bases 1 to 896)
Genoscope.
Direct Submission
Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
segre@genoscope.cns.fr - Web : www.genoscope.cns.fr)
This GSS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.
FEATURES
Location/Qualifiers
1..896
/organism="Saccharomyces servazzii"
/mol_type="genomic DNA"
/strain="CBS 4311"
/db_xref="taxon:27293"
/clone="AT0AA012G10"
/clone_lib="AT0AA"
/note="end : T3"
complement(<282..>892)
/note="similar to Saccharomyces cerevisiae ORF YNL134c [
similarity to C.carbonum toxD gene ]
1 putative frameshift(s)"
/evidence=not experimental
misc_feature
252 a 136 c 190 g 316 t 2 others
BASE COUNT
ORIGIN

Query Match
Best Local Similarity 16.4%; Score 182.4; DB 29; Length 896;
Matches 350; Conservative 1; Mismatches 263; Indels 1; Gaps 1;

```

	Genomic exploration of the hemiascomycetous yeasts: 7.		
TITLE	Saccharomyces servazzii		
JOURNAL	FEBS Lett. 487 (1), 47-51 (2000)		
MEDLINE	20584717		
PUBMED	11152882		
REFERENCE	3 (bases 1 to 881)		
AUTHORS	Genoscope.		
TITLE	Direct Submission		
JOURNAL	Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage, 2 rue Gaston Crémieux, CP 5706, 91057 EVRY cedex, FRANCE. [E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr]		
COMMENT	This GSS is part of a random genomic sequencing program of thirteen yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii, Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila, Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.		
FEATURES			
Source	Location/Qualifiers 1..881 /organism="Saccharomyces servazzii" /mol_type="genomic DNA" /strain="CBS 4311" /db_xref="taxon:27293" /clone="AT0AA012G10" /clone_lib="AT0AA" /note="end : 77"		
misc_feature	<140..>877 /note="similar to Saccharomyces cerevisiae ORF YLR460c [ similarity to C.carbonum toxD protein ]" /evidence=not_experimental		
misc_feature	<143..>877 /note="similar to Saccharomyces cerevisiae ORF YNL134c [ similarity to C.carbonum toxD gene ]" /evidence=not_experimental		
misc_feature	<158..>877 /note="similar to Saccharomyces cerevisiae ORF YCR102c [ similarity to C.carbonum toxD gene ]" /evidence=not_experimental		
BASE COUNT	236 a 231 c 205 g 209 t		
ORIGIN			
Query Match	16.3%; Score 181.2; DB 29; Length 881;		
Best Local Similarity	54.7%; Pred. No. 3.3e-34;		
Matches	404; Conservative 0; Mismatches 328; Indels 6; Gaps 2;		
QY	1 ATGTGAGTTCCAACCACTCAAAGGCCGTTCATTGAAGGTGCACAAAGCTTGTGTAAA 60		
Db	140 ATGTCTTACCACCTCCATCGAGCCACCGTTCATCGACAACGGTAAAGCCGTATCAA 199		
QY	61 ACAGATGTCATGTTCCAGATAAAGAAGGGTACAGCTTGGTGGAAGGTTGAGGCTGTT 120		
Db	200 AC---CGTCCCATTGCCTCAACTGGACGACGAATACGCTTGATCAAGACCAAAGCCGTT 256		
QY	121 GCTGTACCCNACGATTGGAGCATATTCCTTAAGATTGTTGCCAGAGGTTCAATT 180		
Db	257 GCCGGTAACCAACCGATTGGAACATGTCGAAGTTCCGGCCTTAGTCCAAAACGGGGCCGTC 316		
QY	181 CTAGATGTGACATTGCTGGTACAGTTGTCACAACTTGGACCAAAATGCTAGTACTGAC-- 237		
Db	317 ATCGGTGTAGATCGCCCGGTGAATCTCAAGTTAGTTCCAAGTCGACACAAACAA 376		
QY	238 TTGAAGGTTGGAGATACCGGTTTTGGTTTTGTTTTCACGGTGTCTCCCAAACAGATCTCAA 297		
Db	377 TTTCCACGTAGGTGACTATGTCTACGGGTTGCTTTCATGGGCTTCTTAAGAGACGTCCAGAT 436		
QY	298 AATGGTGCATTGCTGATATGCCAGGGTTTATCCACCTTTGTTTTCACAGAGTAACCTTA 357		
Db	437 AATGGGGCTTTGCGAGATACGTTGCTCTAGATTCCATGATCAACCTTCCCTGTCCCCAAC 496		



Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.

```
FEATURES
  source            1..1072
                    /organism="Saccharomyces servazzii"
                    /mol_type="genomic DNA"
                    /strain="CBS 4311"
                    /db_xref="taxon:27293"
                    /clone_lib="XAT0AA001D07"
                    /clone_lib="XAT0AA"
                    /notes="end : T7"
  misc_feature      <336..>1039
                    /note="similar to Saccharomyces cerevisiae ORF YLR460c [
                    similarity to C.carbonum toxD protein ]
                    2 putative frameshift(s)
                    similar to Saccharomyces cerevisiae ORF YNL134c [
                    similarity to C.carbonum toxD gene ]"
                    /evidences=not_experimental
  misc_feature      <348..>1039
                    /note="similar to Saccharomyces cerevisiae ORF YCR102c [
                    similarity to C.carbonum toxD gene ]
                    2 putative frameshift(s)"
                    /evidences=not_experimental
BASE COUNT        350 a 178 c 210 g 331 t 3 others
ORIGIN
Query Match       13.9%; Score 154.6; DB 29; Length 1072;
Best Local Similarity 56.0%; Pred. No. 1.5e-27;
Matches 394; Conservative 1; Mismatches 270; Indels 38; Gaps 4;
QY 8 TTCCAACTCAAAAAGCGCTCATCTATGAAGGTGACAAAGCTGTTGTTAAAACAGATG 67
DB 337 TTCCAAGACTATGAAGCGCTGCTTTTGTATGCACAGTTCGAAAGTTCAAGATAGA 396
QY 68 TCTCAGTTCAGAAATTAAGAGGGGTACAGCTTGGTGAAGTTGAGGCTGTTGCTGGTA 127
DB 397 TTAAGTTCCCTTCGTGGATGATGGTTCCTTTTAAATTCAGTTCTAGCAGTGCACAA 456
QY 128 ACCCACTGATGGACATATGCTTATAAGATTGCTCAGAGGTTCAATTCAGGAT 187
DB 457 ATCCAATGGATGGAGCATGTTGAATTTGAATGGATGGGTCCAAAGAGTCAATCCTAGTT 516
QY 188 GTGACATTCGTGTACAGTTGTCAAACTTGACCA---AATGCTAGTACTGACTTGAAGG 244
DB 517 GTGATGTTGCGGTGATAGTAGTCTATTGGTCAGGGATTCAGCAGATGAATATAAG 576
QY 245 TTGAGATACCGGTTTCGGTTTTGTTCCAGGTGCTTCCCAACAGATCTCTAAAAGTGTG 304
DB 577 TTGATGATAATGTCGTGCAATTTGTTTCATGGGGGTCAATCCGTAACCTGGCAATGGGG 636
QY 305 CATTGCTGATATGCCAGGGTTATCCACCTTGTGTTT-----
DB 637 CATTGACAAATTTTGCTGTAGATGCAATATACATTTAAATTAATTAATTAATTAATTA 696
QY 344 -----ACAAGATAACTTAACTCACTCAACTGCTGATGAATTTTGAAGGCC 391
DB 697 TGCTTCATAGACACACGGCTTTGAAACAGGTCAAGATTTTATCGAATTTGATGAAGGC 756
QY 392 CTGTGAAGAACTTCGAATCTGTCGATCATTTGCCAGTTTCGTTGACAACTGCTGGTGTTA 451
DB 757 CAGTGGATTCATTGAAACTGCTGCTACTATGCTGCTGGCATTAAATCACTGCAGGTGCTA 816
QY 452 GTTTGTCATCACTTCGGCTCAAAAATGGAATGGACACCATCTACCCCGCAACATCTC 511
DB 817 TCTTAGTCCATGAATTTGGGGTTAAATTTAGAGTGGGAMCTGTTTGAACCAACAAATAAT 876
QY 512 ATCCATTATTGATTT--GGGGTGGTGTACAGCAGTGGGTCAACAACTAATCCAGTTGCC 570
DB 877 CACCAAGTTTAAATTTGGGGTGGTGGCAGCAGTGTAGGTCAAAATCTTAATTTCAAGTAAT 936
QY 571 AAACATATCAATGCTTATAAGATTGTAAGTGTGCTTCTTAAAG--CATGAAAGCT 629
```

```
Db 937 AAAAAAATAAATCATCTCTCAAGATTATTACAGTTGCTTCCCAAGACATGAAGATTT 996
QY 630 TTTAAAGTCTTATGGTGTGCTGATGATGCTTTTGACTATCATGAT 672
DB 997 GCTACATCATTTATGGTGCAGATATAATATATGATTATCATGAT 1039

RESULT 5
AZ931755/c
LOCUS             488 bp      DNA      linear      GSS 01-APR-2001
DEFINITION       474.dhz89f04.e1 Saccharomyces unisporus NRRL Y-1556 Saccharomyces
                  unisporus genomic clone 474.dhz89f04.e1, genomic survey sequence.
VERSION          AZ931755
KEYWORDS         GSS.
SOURCE           Saccharomyces unisporus
ORGANISM         Saccharomyces unisporus
                  Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
                  Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
REFERENCE        1 (bases 1 to 488)
AUTHORS          Clifton,P.F., Hillier,L.W., Fulton,L., Graves,T., Miner,T., Gish
                  W.R., Waterston,R.H. and Johnston,M.
TITLE            Surveying Saccharomyces genomes to identify functional elements by
                  comparative DNA sequence analysis
JOURNAL          Unpublished
COMMENT          Contact: Johnston M
                  Department of Genetics
                  Washington University Medical School
                  Box 8232, 4566 Scott Ave., St. Louis, MO 63110, USA
                  Tel: 314 362 2735
                  Fax: 314 362 7855
                  Email: mj@genetics.wustl.edu
                  Class: random plasmid subclone.
FEATURES         Location/Qualifiers
                  1..488
                   /organism="Saccharomyces unisporus"
                   /mol_type="genomic DNA"
                   /strain="NRRL Y-1556 (CBS 398)"
                   /db_xref="taxon:27294"
                   /clone="474.dhz89f04.e1"
                   /clone_lib="Saccharomyces unisporus NRRL Y-1556"
                   /notes="Random genomic sequence"
BASE COUNT       154 a 97 c 79 g 158 t
ORIGIN
Query Match       13.2%; Score 146.6; DB 28; Length 488;
Best Local Similarity 58.2%; Pred. No. 1.1e-25;
Matches 279; Conservative 0; Mismatches 194; Indels 6; Gaps 1;
QY 256 GGTTCGTTTCTGTTTCACGTCGTTCCCAACAGATCCTTAAATAATGTCATTTGCTGAA 315
DB 482 GTTCTGCTATTGTTAATGGTGTGCTGCACTGAATTCAGATTAATGGCGTTTTCGCTGAA 423
QY 316 TATGCCAGGGTTTATCCACCTTTGTTTTCACAG-----AGTAACCTTAACCTCACTCAACT 369
DB 422 TACGTGAAGTGGATTCGAATACGTCTTCAAGTTCACCAAAAATTAATTTCTTCTAAC 363
QY 370 GCTGATGAATTTCTGAAGCCCTCTGAAGAACTTCGAATCTGCTGTCATCAITGCCAGTT 429
DB 362 AAGAATAATATCGAGGAGGTAATTTGACACATTCGGGAAGCTTGTGTAGTTTCCCAATG 303
QY 430 TCCTTGACAACTGCTGCTGTTAGTTTGTGTCATCACTTGGGCTCAAAAATGGAATGGCAC 489
DB 302 GTATTGTATACCGCTTTAGCTGCTGTTTCTACGAATTAATAATTTGAATTTGAATGGCAA 243
QY 490 CCATACCCCGCAACATACCTCATCTCATTTTTCATTTGGGGTGGTGTGTACAGCAGTGGGT 549
DB 242 CCAGAAAAACCAACATGATCATCAATCTTAATTTGGGGTGGTGTGTACTGCTATTTGGT 183
QY 550 CAACAACTAATCAAGTTGCCAAACATATCAATGCTTTATACTAAGATTGTGAAGTTGCT 609
DB 182 CAATATGCTATTCAAATTCGTCAGCAAGTTTCATGGATTTTACCAGAAATTTATTACAGTGGCT 123
```

```

610 TCTAAAGGATGAAAGCTTTTAAAGTCTTATGCTGCTGATGCTTTTGGATCAT 669
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
122 TCAAGAAACATGAAGAGCTTAAATCTTTGGTGCAGATGAAGTTCGATTATCAT 63
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
670 GATCAGCGGTTATTGAGCAGATCAATTCGAAGTATCCAAACCTGCAACATGTTATTGA 728
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
62 GACAAGGATGTCGTGGATCAATTTACCAAGAAATATGATAATAAAGAACCTTAATTGA 4
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 6
CNS06LW5/c
LOCUS
DEFINITION
T3 end of clone XAT0AA002A04 of library XAT0AA from strain CBS 4311
of Saccharomyces servazzii, genomic survey sequence.
ACCESSION
AL404715
VERSION
AL404715.1 GI:12166529
KEYWORDS
GSS.
SOURCE
Saccharomyces servazzii
ORGANISM
Saccharomycetes servazzii
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
REFERENCE
1 (bases 1 to 1118)
Souciet, J.L., Aigle, M., Artiguenave, F., Blandin, G.,
Bolotin-Fukuhara, M., Bon, E., Brottier, P., Casaregola, S.,
de-Montigny, J., Dujon, B., Durrens, P., Lepingle, A., Llorente, B.,
Malpertuy, A., Neuveglise, C., Ozier-Kalogeropoulos, O., Potier, S.,
Saurin, W., Tekaita, F., Toffano-Nioche, C., Wesolowski-Louvel, M.,
Wincker, P. and Weissenbach, J.
Genomic exploration of the hemiascomycetous yeasts: 1. A set of
yeast species for molecular evolution studies
PFBs Lett. 487 (1), 3-12 (2000)
20584711
11152876
REFERENCE
2 (bases 1 to 1118)
Casaregola, S., Lepingle, A., Bon, E., Neuveglise, C., Nguyen, H.,
Artiguenave, F., Wincker, P. and Gaillardin, C.
Genomic exploration of the hemiascomycetous yeasts: 7.
Saccharomycetes servazzii
PFBs Lett. 487 (1), 47-51 (2000)
20584717
11152882
REFERENCE
3 (bases 1 to 1118)
Genoscope.
Direct Submission
Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
segref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
This GSS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.
Location/Qualifiers
1. .1118
/organism="Saccharomyces servazzii"
/mol_type="genomic DNA"
/strain="CBS 4311"
/db_xref="taxon:27293"
/clone="XAT0AA002A04"
/clone_lib="XAT0AA"
/note="end : T3"
complement(<136. .>1107)
/note="similar to Saccharomyces cerevisiae ORF YNL134c [
similarity to C.carbonum toxD gene ]
2 putative frameshift(s)"
/evidence=not experimental
complement(<136. .>1104)

misc_feature
/note="similar to Saccharomycetes cerevisiae ORF YCR102c [
similarity to C.carbonum toxD gene ]
2 putative frameshift(s)"
/evidence=not experimental
328 t
5 others

BASE COUNT 366 a 217 c 202 g 328 t 5 others
ORIGIN
Query Match 12.3%; Score 137.2; DB 29; Length 1118;
Best Local Similarity 54.1%; Pred. No. 3.3e-23;
Matches 379; Conservative 1; Mismatches 294; Indels 26; Gaps 4;

QY 11 CAACCACTCAAAAGCCGTCATCATTTGAAGGTGACAAAGCTGTTGTTAAACAGATCTCT 70
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1112 CAACCTCTATGAAGCCACCGTTATTCGACAAATGTAATCTCTATCAAAAC---TGTGC 1056
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 71 CAGTTCAGAAATTAAGAGGGGTACAGCCTTGGTGAAGGTGAGGCTGTTGCTGGTAACC 130
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1055 CATTGCCCAACTGGACGACGATACGCTCTTACCCAGGACCAATCAGTTGCTCGTAATC 996
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 131 CAACTGATTGGAGCATATTTGCTTATAAGATTGGTCCAGAAGCTTCAATTTCTAGGATGTG 190
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 995 CGACGGGTTGAACACCTAGCAAGTATGCTCTTGGGCCAATTTGGTCCGTGATTGGTGTG 936
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 191 ACATTGCTGTGTACAGTTGTCAAACTTTGGACCAAAATG---CTAGTACTGACTTGAAGGTG 247
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 935 ATGCCGCTGTGAAATTTGTAATTTAGTCTTAAAGTCTATACACTAAATCCACCTAG 876
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 248 GAGATACCGGTTTCGGTGTGTTGTTTCACGGTGTCTTCCAAACAGATCTTAAATATGTTGCAT 307
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 875 GTGACTGTGTTATGTTATTTATTCATGCGCTCTTCCAGAGACGCTCTCGACACGGTGCAT 816
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 308 TTGCTGAATATGCCAGGGTTTATCCACCTTTGTTTACAGAGTAACTTAACTCACTCAA 367
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 815 TTGCCAGTACGCTGCTGTGATTTCTTAAGATCTCATTTATA----- 774
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 368 CTGCTGATGAAATTTCTGAAGGCCCTGTGAAGAACTTCGAAATCTGCTGCTCATTTGCCAG 427
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 773 -TGAATAAGGGGTTTTTCAGTGAAGATTTTCACTCCCTGGAAAGGTGCTGCTGCTATCCC 715
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 428 TTTGTTGACACTGCTGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 487
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 714 CGTCTTTGTTGACAGCTCTCTGCAACTCTTTTCT-ACACGTGCTTTAAAAAGGAATGGC 656
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 488 ACCCATCTACCCCGCAACATACATCATTCATTATTGTTGTTGGGTGTTGTTGTTGTTGTTGTTG 547
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 655 AACCAAGAGAGGCCCAAAATAAAGGTACTGTTTTTMAITTTGGATGTTGTTGTTGTTGTTGTTG 596
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 548 GTCAACAACTTAATCCAAAGTTGCAAAACATATCAATGCTTATATAAGATTGTAACCTGTTG 607
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 595 CTCTAGCCATTTTACAAATTTTCCAAAAATTTCAATGTTGCTGAGAAATTTGTTGCTGTTG 536
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 608 CTTCTAAAAGCATGAAAAGCTTTTAAAGCTTTATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 667
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 535 CTTCTGAAGAAACGTTGAAACTTAATCTTAAGAAATTTGGGTGTTGTTGTTGTTGTTGTTGTTG 476
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 668 ATGATCAGCGGCTTATTGAGCAGATCAATTCGAAGTATCC 707
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 475 ATGATGCTGATATTGTTGAACAAATCAGTTCCAAATATATAC 436
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 7
LOCUS
AZ9311754
DEFINITION
474.dhz89f03.s1 Saccharomycetes unisporus NRRL Y-1556 Saccharomycetes
unisporus genomic clone 474.dhz89f03.s1, genomic survey sequence.
ACCESSION
AZ9311754
VERSION
AZ9311754.1 GI:13502665
KEYWORDS
GSS.
SOURCE
Saccharomycetes unisporus
ORGANISM
Saccharomycetes unisporus
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
REFERENCE
1 (bases 1 to 434)

```

**AUTHORS** Clifton, P.F., Hillier, L.W., Fulton, L., Graves, T., Miner, T., Gish, W.R., Waterston, R.H. and Johnston, M.

**TITLE** Surveying Saccharomyces genomes to identify functional elements by comparative DNA sequence analysis

**JOURNAL** Unpublished

**COMMENT** Contact: Johnston M  
Department of Genetics  
Washington University Medical School  
Box 8232, 4566 Scott Ave., St. Louis, MO 63110, USA  
Tel: 314 362 2735  
Fax: 314 362 7855  
Email: mj@genetics.wustl.edu  
Class: random plasmid subclone.  
Location/Qualifiers  
1. .434  
/organism="Saccharomyces unisporus"  
/mol\_type="genomic DNA"  
/strain="NRRL Y-1556 (CBS 398)"  
/db\_xref="taxon:27294"  
/clone="474.dhz89f03.s1"  
/clone\_lib="Saccharomyces unisporus NRRL Y-1556"  
/note="Random genomic sequence"

**BASE COUNT** 133 a 76 c 88 g 137 t

**ORIGIN**

Query Match 11.9%; Score 132.6; DB 28; Length 434;  
Best Local Similarity 59.7%; Pred. No. 3.4e-22;  
Matches 244; Conservative 0; Mismatches 159; Indels 6; Gaps 1;

256 GGTTCGGTTTGTTCACGGTCTCCCAACAGATCTTAAATGTGTGCTGAA 315  
|||||  
26 GGTTCGGTGTGATTAATGGTGTATGTTCAATTCAGATATGCTTATCTGAA 85  
|||||  
316 TATGCCAGGGTTTATCCACCTTGTGTTTACAA-----GAGTACTTAATCACTCACT 369  
|||||  
86 TATGTTAAGTGGACCAACACTATTTTCAATTTCCCAAGACATTAATTTTACTAAA 145  
|||||  
370 GCTGATGAATTTCTGAAGGCCCTGTGAAGAACTTCGAATCTGCTGCATCATGCCAGTT 429  
|||||  
146 AAGACTGACGTGGCAACAGCAAGTTCGACACATGGGAAGCTTGTGCCAGTTCCCAATG 205  
|||||  
430 TCGTTGCAACTGCTGTGTAGTTGTGTCATCACTTGGGCTCAAAATAGGAATGGCAC 489  
|||||  
206 GTGTTCTACACTGCTTAGCTGTACCTTTTACCAATTTGAAATTTAAATTTAGATGCAA 265  
|||||  
490 CCATCTACCCGCAACATACTCACTTATTTGATTTGGGGTGTGTACAGCAGTGGT 549  
|||||  
266 CCAGAAGGCCACAAATATGACCATCAATCTTAATTTGGGGTGTGTCCACTGGTGTGGT 325  
|||||  
550 CAACAACATAATCCAAAGTTGGCAACATATCAATGCTTATTAAGATTTGAATGTTGCT 609  
|||||  
326 CAATATGCTATCAATTTGTTAAACAGTCCATGGTTTACTAGATATTGTTGTTGCC 385  
|||||  
610 TCTAAAAGCATGAAAGCTTTTAAAGTCTTATGGTGTGCTGATGATGCT 658  
|||||  
386 TCTAAGAAACATGAAAGTTGTTAAATCTTATGCTGTGATGAAATCT 434  
|||||

**FEATURES**  
source  
1. .454  
/organism="Saccharomyces cerevisiae"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:4932"  
/lab\_host="E. coli"  
/clone\_lib="mTn-3xHA/lacZ Insertion Library"  
/note="Vector: pHS56-Sal; A yeast genomic DNA library (lacking mitochondrial DNA) was prepared in pHS56-Sal; genomic DNA was size-fractionated (DNA of roughly 2-3 kb in length) prior to cloning. This library was subsequently mutagenized with a mTn-3xHA/lacZ minitransposon containing lacZ, URA3, and tet resistance."

**BASE COUNT** 122 a 106 c 102 g 124 t

**ORIGIN**

Query Match 11.4%; Score 126.6; DB 28; Length 454;  
Best Local Similarity 64.7%; Pred. No. 1.1e-20;  
Matches 205; Conservative 0; Mismatches 109; Indels 3; Gaps 1;

4 TCAGTTCACCACTCAAAAGCCGTCATTCATTGAAGGTGACAAAGCTGTGTAAACA 63  
|||||  
401 TCGATTCCAGAAACCATGAAGCCGTTGTCATTGAAATGGCAAGCGTGTAGTCAACAG 342  
|||||  
64 GATGCTCTCAGTTCAGCAATTAAGAGGAGGTACAGCCTTGGTGAAGGTTGAGGCTGTGCT 123  
|||||  
341 GACATTCCAATTCCTGAATTTAGAAAGAGGATTTGTTCTAATTAAGACTGTGCGGTTGCC 282  
|||||  
124 GGTAAACCACTGATTTGGAAGCATATTTGCTTATAAGATTGGTCCAGAAGGTTCAATCTA 183  
|||||  
281 GGTAAACCTACCATTTGGAACATATTTGATTTCAAGATTGGTCTCTCAAGGTGCTCTTA 222  
|||||  
184 GGATGTGACATTCGCTGTGACAGTTGTCAAACTTTGGACCAAAATGCTAGTACTG---ACTTG 240  
|||||  
221 GGCTGTGATGTCAGCCGCGCAAAATCGTAAAGTTGGGCCCAAAATGTTGATGCTGCACGCTTT 162  
|||||  
241 AAGGTTGGAGATACCGGTTTCGGTTTGTTCACGGTGTCTCCCAACAGATCTCAAAAT 300  
|||||  
161 GCCATTGGGATGATACATTTATGGGTTATTTCACGGTGTCTCAGTGAGGTTCCCTCAAC 102  
|||||  
301 GGTGCAATTTGCTGAATA 317  
|||||  
101 GGTGCTCTTGTGCTGAGTA 85  
|||||

**RESULT 9**  
AQ01781/c  
LOCUS  
DEFINITION V7483 mTn-3xHA/lacZ Insertion Library, strain AB972 Saccharomyces cerevisiae genomic 5', genomic survey sequence.  
ACCESSION AQ01781  
VERSION AQ01781.1  
KEYWORDS GSS.  
SOURCE Saccharomyces cerevisiae (baker's yeast)  
ORGANISM Saccharomyces cerevisiae  
1 (bases 1 to 454)  
REFERENCE Ros-Macdonald, P., Roemer, T., Coelho, P.S.R., Agarwal, S., Kumar, A., deStages, S.A., Cheung, K.-H., Sheehan, A., Symoniat, D., Jansen, R.,

**REFERENCE** 1 (bases 1 to 546)

**AUTHORS** Umansky, L., Heidtman, M., Nelson, K., Iwasaki, H., Kanada, D., Lugo, R., Hager, K., Miller, P., Roeder, G.S. and Snyder, M.

**TITLE** Large-Scale Analysis of the Yeast Genome by Transposon Tagging and Gene Disruption

**JOURNAL** Unpublished

**COMMENT** Contact: Kumar A  
Dept. of Mol. Cell. and Dev. Biology  
Michael Snyder  
Yale University  
P.O. Box 208103, New Haven, CT 06520-8103, USA  
Tel: 203 432 9949  
Fax: 203 432 6161  
Email: anuj.kumar@yale.edu  
te of mTn-3xHA/lacZ insertion.  
Seq primer: GGCCTTCCTTTCTTTGGAAGTAC  
Class: transposon-tagged.  
Location/Qualifiers  
1. .454  
/organism="Saccharomyces cerevisiae"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:4932"  
/lab\_host="E. coli"  
/clone\_lib="mTn-3xHA/lacZ Insertion Library"  
/note="Vector: pHS56-Sal; A yeast genomic DNA library (lacking mitochondrial DNA) was prepared in pHS56-Sal; genomic DNA was size-fractionated (DNA of roughly 2-3 kb in length) prior to cloning. This library was subsequently mutagenized with a mTn-3xHA/lacZ minitransposon containing lacZ, URA3, and tet resistance."

**BASE COUNT** 122 a 106 c 102 g 124 t

**ORIGIN**

Query Match 11.4%; Score 126.6; DB 28; Length 454;  
Best Local Similarity 64.7%; Pred. No. 1.1e-20;  
Matches 205; Conservative 0; Mismatches 109; Indels 3; Gaps 1;

4 TCAGTTCACCACTCAAAAGCCGTCATTCATTGAAGGTGACAAAGCTGTGTAAACA 63  
|||||  
401 TCGATTCCAGAAACCATGAAGCCGTTGTCATTGAAATGGCAAGCGTGTAGTCAACAG 342  
|||||  
64 GATGCTCTCAGTTCAGCAATTAAGAGGAGGTACAGCCTTGGTGAAGGTTGAGGCTGTGCT 123  
|||||  
341 GACATTCCAATTCCTGAATTTAGAAAGAGGATTTGTTCTAATTAAGACTGTGCGGTTGCC 282  
|||||  
124 GGTAAACCACTGATTTGGAAGCATATTTGCTTATAAGATTGGTCCAGAAGGTTCAATCTA 183  
|||||  
281 GGTAAACCTACCATTTGGAACATATTTGATTTCAAGATTGGTCTCTCAAGGTGCTCTTA 222  
|||||  
184 GGATGTGACATTCGCTGTGACAGTTGTCAAACTTTGGACCAAAATGCTAGTACTG---ACTTG 240  
|||||  
221 GGCTGTGATGTCAGCCGCGCAAAATCGTAAAGTTGGGCCCAAAATGTTGATGCTGCACGCTTT 162  
|||||  
241 AAGGTTGGAGATACCGGTTTCGGTTTGTTCACGGTGTCTCCCAACAGATCTCAAAAT 300  
|||||  
161 GCCATTGGGATGATACATTTATGGGTTATTTCACGGTGTCTCAGTGAGGTTCCCTCAAC 102  
|||||  
301 GGTGCAATTTGCTGAATA 317  
|||||  
101 GGTGCTCTTGTGCTGAGTA 85  
|||||

**RESULT 9**  
AQ01781/c  
LOCUS  
DEFINITION V7483 mTn-3xHA/lacZ Insertion Library, strain AB972 Saccharomyces cerevisiae genomic 5', genomic survey sequence.  
ACCESSION AQ01781  
VERSION AQ01781.1  
KEYWORDS GSS.  
SOURCE Saccharomyces cerevisiae (baker's yeast)  
ORGANISM Saccharomyces cerevisiae  
1 (bases 1 to 546)  
REFERENCE Ros-Macdonald, P., Roemer, T., Coelho, P.S.R., Agarwal, S., Kumar, A., deStages, S.A., Cheung, K.-H., Sheehan, A., Symoniat, D., Jansen, R.,

AUTHORS	Ross-Macdonald, P., Roemer, T., Coelho, P. S. R., Agarwal, S., Kumar, A., desErgues, S. A., Cheung, K. -H., Sheehan, A., Symoniatis, D., Jansen, R., Unanue, B., Heidtman, M., Nelson, K., Iwasaki, H., Kanada, D., Lugo, R., Hager, K., Miller, P., Roeder, G. S. and Snyder, M.
TITLE	Large-Scale Analysis of the Yeast Genome by Transposon Tagging and Gene Disruption
JOURNAL	Unpublished
COMMENT	Contact: Kumar A Michael Snyder, Dept. of Mol. Cell. and Dev. Biology Yale University P.O. Box 208103, New Haven, CT 06520-8103, USA Tel: 203 432 9949 Fax: 203 432 6161 Email: anuj.kumar@yale.edu te of min-3XHA/lacZ insertion. Seq primer: GGCTCTCTTTTGGAGTAC Class: transposon-tagged.

BASE COUNT	157 a	121 c	114 g	150 t	4 others
ORIGIN					
location/qualifiers					
source					
	1. 546				
	/organism="Saccharomyces cerevisiae"				
	/mol_type="genomic DNA"				
	/strain="AB972 - trp1 r(0) (S288C background)"				
	/db_xref="taxon:4932"				
	/lab_host="E. coli"				
	/clone_lib="mtm-3xHA/lac2 Insertion Library, strain AB972"				
	/note="vector: PHSS6-Sal; A yeast genomic DNA library was prepared in PHSS6-Sal; genomic DNA was size-fractionated (DNA of roughly 2-3 kb in length) prior to cloning. This library was subsequently mutagenized with a mtm-3xHA/lac2 minitransposon containing lacZ, URA3, and tet resistance."				

RESULT 10	
AQ873797/c	
LOCUS	537 bp DNA linear
DEFINITION	W75E1 mTn-3XHA/lacZ Insertion Library, strain AB972 <i>Saccharomyces cerevisiae</i> genomic 5', genomic survey sequence.
ACCESSION	AQ873797
VERSION	AQ873797.1
KEYWORDS	GI:6286041
SOURCE	GSS.
ORGANISM	<i>Saccharomyces cerevisiae</i> (baker's yeast)
	<i>Saccharomyces cerevisiae</i>

Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.	
1 (bases 1 to 537)	
Ross-Macdonald, P., Roemer, T., Coelho, P. S. R., Agarwal, S., Kumar, A., deStagosa, S. A., Cheung, K. -H., Sheehan, A., Symoniatig, D., Jansen, R., Umansky, L., Heidtman, M., Nelson, K., Iwasaki, H., Kanada, D., Lugo, R., Hager, K., Miller, P., Roeder, G. S. and Snyder, M.	
Large-Scale Analysis of the Yeast Genome by Transposon Tagging and Gene Disruption	
Unpublished	
Contact: Kumar A	
Michael Snyder, Dept. of Mol. Cell. and Dev. Biology	
Yale University	
P.O. Box 208103, New Haven, CT 06520-8103, USA	
Tel: 203 432 9949	
Fax: 203 432 6161	
Email: anuj.kumara@yale.edu	

```

FEATURES
    source
        1..537
            /organism="Saccharomyces cerevisiae"
            /mol_type="genomic DNA"
            /strain="AB972 - trp1 r(0) (S288C background) "
            /db_xref="taxon:4932"
            /lab_host="E. coli"
            /clone_lib="mTn-3xHA/lacZ Insertion Library, strain AB972"
            /note="Vector: pHS6-Sal; A yeast genomic DNA library was
            prepared in pHS6-Sal; genomic DNA was size-fractionated
            (DNA of roughly 2-3 kb in length) prior to cloning. This
            library was subsequently mutagenized with a mTn-3xHA/lacZ
            mini transposon containing lacZ umu3 and recombination
            sites for transposon excision."
    Seq primer: GGCTTCCTCTTCCTTGGAGTAC
    Class: transposon-tagged
    Location/Qualifiers

```

Db	115	GGTGCCTTTGCTGAGTA	99
RESULT 11			
AQ502797/c			
LOCUS	AQ502797	529 bp	DNA
DEFINITION	linear GSS 29-APR-1999		
	V75D10 mtN-3xHA/lacZ Insertion Library <i>Saccharomyces cerevisiae</i>		
	genomic 5', genomic survey sequence.		
ACCESSION	AQ502797		
VERSION	AQ502797.1	GI:4705343	

KEYWORDS Saccharomyces cerevisiae (baker's yeast)

SOURCE Saccharomyces cerevisiae

ORGANISM Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomycetes; 1 (bases 1 to 529)

REFERENCE Ross-Macdonald, P., Roemer, T., Coelho, P.S.R., Agarwal, S., Kumar, A., deBettages, S.A., Cheung, K.-H., Sheehan, A., Symoniat, D., Jansen, R., Umansky, L., Heidtman, M., Nelson, K., Iwasaki, H., Kanada, D., Lugo, R., Hager, K., Miller, P., Roeder, G.S. and Snyder, M.

AUTHORS Large-Scale Analysis of the Yeast Genome by Transposon Tagging and Gene Disruption

TITLE Unpublished

JOURNAL Contact: Kumar A

COMMENT Michael Snyder, Dept. of Mol. Cell. and Dev. Biology  
Yale University  
P.O. Box 208103, New Haven, CT 06520-8103, USA  
Tel: 203 432 9949  
Fax: 203 432 6161  
Email: anuj.kumar@yale.edu  
te of mTn-3xHA/lacZ insertion.  
Seq primer: GGCCTCTCTTCTTGGAGTAC  
Class: transposon-tagged.

FEATURES source

Location/Qualifiers

1..529

/organism="Saccharomyces cerevisiae"

/mol\_type="genomic DNA"

/db\_xref="taxon:4932"

/lab\_hosts="E. coli"

/clone\_lib="mTn-3xHA/lacZ Insertion Library"

/notes="Vector: pHS56-Sal; A yeast genomic DNA library (lacking mitochondrial DNA) was prepared in pHS56-Sal; genomic DNA was size-fractionated (DNA of roughly 2-3 kb in length) prior to cloning. This library was subsequently mutagenized with a mTn-3xHA/lacZ minitransposon containing lacZ, URA3, and tet resistance."

BASE COUNT 152 a 117 c 112 g 147 t 1 others

Query Match 11.1%; Score 124; DB 28; Length 529;  
Best Local Similarity 64.0%; Pred. No. 5.1e-20;  
Matches 203; Conservative 0; Mismatches 111; Indels 3; Gaps 1;

Qy 4 TCAGTTCACCACTCAAAAGCCGTCATCATTAAGAGTGACAAAGCTGTGTTAAACA 63

Db 415 TCGATTCCAGAAACCATGAAAGCGTTGTCATTGAAATGCAAGGCTGTAGTCANACAG 356

Qy 64 GATGCTCAGTTCAGAAATTAAGAGGGTACAGCCTTGGTGAAGGTTGAGGCTGTGCT 123

Db 355 GACATTCCAATTCCTGAATTAGAGAAGGATTTGTTCTAATTAAGACTGTGCGCGTAGCC 296

Qy 124 GGTAAACCACTGATTCGAGCATATTGCTTATAGATTGGTCCAGAGGTTCAATTCTA 183

Db 295 GGTAAACCTCCGATTCGAAACATATTGATTTCAAGATTGTTCTCAAGGTCCTCTTA 236

Qy 184 GGATGTGACATTTGCTGTACAGTTGTCAAACTGGACCAATGCTAGTACTG---ACTTG 240

Db 235 GGCTGTGATGACGCGGCCAAATCGTAAGTTGGCCCAATGTTGATGCTGACGCTTT 176

Qy 241 AAGTTGGAGATACCGGTTTCGGTTTGTTCACGGTGTCTCCCAACAGATCTTAAANAAT 300

Db 175 GCCATTGGTATTACATTTATGGGGTTATTACCGTGTCTTACCGTGTCTTCCCTCAAAC 116

Qy 301 GGTCATTTGCTGAATA 317

Db 115 GGTGCTTTGCTGAGTA 99

RESULT 12

LOCUS BZ300244/c 467 bp DNA linear GSS 31-OCT-2002

DEFINITION KD0613.R1 Kluyveromyces delphensis Random Genomic Library  
Kluyveromyces delphensis genomic clone KD0613, genomic survey

sequence.

ACCESSION BZ300244

VERSION BZ300244.1 GI:24444834

KEYWORDS GSS.

SOURCE Kluyveromyces delphensis

ORGANISM Kluyveromyces delphensis

REFERENCE Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Kluyveromyces. 1 (bases 1 to 467)

AUTHORS Wong, S., Fares, M.A., Zimmermann, W., Butler, G. and Wolfe, K.H.

TITLE Evidence from comparative genomics for a complete sexual cycle in the 'asexual' pathogenic yeast *Candida glabrata*

JOURNAL Genome Biol. 4 (2), R10 (2003)

COMMENT MEDLINE 22508158

PUBMED 12620120

CONTACT: Wong S

Department of Genetics, Smurfit Institute

Trinity College Dublin

Dublin 2, Ireland

Tel: 353 1 6082319

Fax: 353 1 6798558

Email: swong@tcd.ie

Class: plasmid ends.

FEATURES source

Location/Qualifiers

1..467

/organism="Kluyveromyces delphensis"

/mol\_type="genomic DNA"

/strain="CBS 2170"

/db\_xref="taxon:51657"

/clone\_lib="KD0613"

/clone\_lib="Kluyveromyces delphensis Random Genomic Library"

BASE COUNT 118 a 108 c 118 g 123 t

Query Match 11.1%; Score 123.8; DB 29; Length 467;  
Best Local Similarity 57.8%; Pred. No. 5.5e-20;  
Matches 262; Conservative 0; Mismatches 182; Indels 9; Gaps 2;

Qy 60 AACAGATGCTCAGTTCAGAAATTAAGAGGGGTACAGCCTTGGTGAAGGTTGAGGCTGT 119

Db 461 AAAGGAGGTTCTTTCGACCAATGGGTGAAACCACTGTCTGATTAAGAACACTGCGGT 402

Qy 120 TGCTGGTAAACCCCACTGATTGGAAGCATATTGCTTAAGATTGGTCCAGAGGTTCAAT 179

Db 401 TGCTGTTAAACCACTGACTGGAGGACATCGCATACAGATCGGTCCACAGGATCAT 342

Qy 180 TCTAGGATGTGACATTTGCTGTACAGTTGTCAAACTTTGGACCAAAAT---GCTAGTACTGA 236

Db 341 CTGTGGCTGTGACGCTACTGCTGAGATTGTCAAGTTAGGTGCTGGTGTCTTCCAA 282

Qy 237 CTTGAAGGTTGGAGATACCGGTTTCGGTTTGTTCACGGTGTCTCCCAACAGATCTCAA 296

Db 281 GTACAAGCTCGGTCAAAAGGTCTACGGTTTGGTCCACGGTTCCTCCGCTCAAGCACCTCA 222

Qy 297 AAATGGTGCAATTTGCTGAATATCCAGGGTTTATCCACCTTTGTTTATC-----AAGAG 350

Db 221 AAACGGTGCCTTCGAGAAATTTCTGTTTGGATTCCTAAAGTCATGTACGTTCCAAACCA 162

Qy 351 TAACTTAACCTCACTCAACTGCTGATGAAATTTCTGAAGGCCCTGTGGAAGAACTTCGAATC 410

Db 161 AGATATTATCCATCAGCGGTAAGGACGAGATCCCAGCTGGTCCAGTTAAGCATGTTGAAGA 102

Qy 411 TGCTGCATCATTCGCCAGTTTCGTTTGACAACTGCTGGTGTAGTTTGTGTCATCACTGGG 470

Db 101 CACAGCAACATTCCTCAGTCTCGTTTGACCACTGCTGGTGTACTGTTTTCACCACTTTCGG 42

Qy 471 CTCAAAAATGGAATGGCACCCCATCTACCCGCA 503

Db 41 TCTAGATTGCTTGGGAACCATCCAGGCCACA 9

RESULT 13

CNS06HG8 946 bp DNA linear GSS 30-NOV-2001  
 LOCUS T7 end of clone AS0AA012C02 of library AS0AA from strain CLIB 533  
 DEFINITION of Saccharomyces bayanus, genomic survey sequence.

ACCESSION AL398958  
 VERSION AL398958.1 GI:12152890  
 KEYWORDS GSS.

SOURCE  
 ORGANISM

Saccharomyces bayanus  
 Saccharomyces bayanus  
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 Saccharomycetales; Saccharomycetaceae; Saccharomyces.

REFERENCE  
 AUTHORS

1 (bases 1 to 946)  
 Souciot, J.L., Aigle, M., Artiguenave, F., Blandin, G.,  
 Bolotin-Fukuhara, M., Bon, E., Brottier, P., Casaregola, S.,  
 de-Montigny, J., Dujon, B., Durrens, P., Lepingle, A., Llorente, B.,  
 Malpertuy, A., Neuveglise, C., Ozier-Kalogeropoulos, O., Potier, S.,  
 Saurin, W., Tekala, F., Toffano-Nioche, C., Wesolowski-Louvel, M.,  
 Wincker, P. and Weissenbach, J.

TITLE  
 Genomic exploration of the hemiascomycetous yeasts: 1. A set of  
 yeast species for molecular evolution studies

JOURNAL  
 MEDLINE

FEBS Lett. 487 (1), 3-12 (2000)  
 20584711  
 11152876

REFERENCE  
 AUTHORS

2 (bases 1 to 946)  
 Bon, E., Neuveglise, C., Casaregola, S., Artiguenave, F., Wincker, P.,  
 Aigle, M. and Durrens, P.

TITLE  
 JOURNAL

Medline  
 MEDLINE  
 20584715  
 11152880

REFERENCE  
 PUBLISHED

3 (bases 1 to 946)  
 Genoscope.  
 Direct Submission

COMMENT

Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,  
 2 rue Gaston Cremieux, CP 706, 91057 EVRY cedex, FRANCE. (E-mail :  
 segre@genoscope.cns.fr - Web : www.genoscope.cns.fr)  
 This GSS is part of a random genomic sequencing program of thirteen  
 yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces  
 exiguus, Saccharomyces servazii, Zygosaccharomyces rouxii,  
 Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces  
 lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia  
 angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,  
 Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to  
 5 kb were prepared and both extremities were sequenced. See  
 keywords for description of this sequence and for the sequence of  
 the other extremity of this insert.

FEATURES  
 source

1. 946  
 /organism="Saccharomyces bayanus"  
 /mol\_type="genomic DNA"  
 /strain="CLIB 533"  
 /variety="uvarum"  
 /db\_xref="taxon:4931"  
 /clone="AS0AA012C02"  
 /clone\_lib="AS0AA"  
 /note="end : T7"  
 <485. .>944  
 /note="similar to Saccharomyces cerevisiae ORF YNL134c [  
 similarity to C.carbonum toxD gene ]  
 2 putative frameshift(s)"  
 /evidence=not\_experimental  
 <523. .>944  
 /note="similar to Saccharomyces cerevisiae ORF YKL102c [  
 similarity to C.carbonum toxD gene ]  
 1 putative frameshift(s)  
 similar to Saccharomyces cerevisiae ORF YLR460c [  
 similarity to C.carbonum toxD protein ]"  
 /evidence=not\_experimental

misc\_feature

252 a 219 c 200 g 271 t 4 others  
 10.7%; Score 119; DB 29; Length 946;

misc\_feature

252 a 219 c 200 g 271 t 4 others  
 10.7%; Score 119; DB 29; Length 946;

BASE COUNT  
 ORIGIN

Query Match

Best Local Similarity 59.8%; Pred. No. 1.1e-18;  
 Matches 253; Conservative 1; Mismatches 161; Indels 8; Gaps 3;

Qy 32 TCATTGAAGTGAACAAGCTGTTGTTAAACACATGCTCAGTTCAGAGTAATTAAGGAGG 91  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 524 TTATTGAAGCGGATGAAGCTGTCTCATCAAGGAGGGTGTGCCCATTCGCGAATTAGAGGATG 583  
 Qy 92 GTACAGCCTTGGTGAAGGTTGAGGCTGTTGCTGTAACCCAACTGATTGGAAGCATATTG 151  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 584 GTTTCGTTCTGATCAAAACCGTCTGTTGCGGTAATCCCACTGATTGGGCACATATTG 643  
 Qy 152 CTTATAAGATTTGTTCCAGAAGTTCAATTCTAGGATGTGACATTTGCTGGTACAGTTGTCA 211  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 644 ACTACAAGATAGGCCCTCAAGGTTCTATCTCTGGTGGCGATGCTACAGGTCCAATTGTTA 703  
 Qy 212 AACTTGGACCAAT---GCTAGTACTGACTTGAAGTTGGAGATACCGGTTTCGGTTTGG 268  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 704 AACTGGGTCGGGTGTGATGTCGATGGCAACGCTTTCCTCCATTGGAGATTACATTTACGGGTTA 763  
 Qy 269 TTCACGGTGTCTCCCAACACAGATCCTAAATAATGCTGTCATTTGCTGAATATGCCAGGTTT 328  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 764 TTCACGGTGTCTCCGTTGAGATTCCCTCCAAUGTGCTTCGCTGATATCTGCAATTC 823  
 Qy 329 ATCCACCTTTGTTTATCAA---GAGTAACCTTAACCTCACTCAACTGCTGATGAAA--TTTC 383  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 824 CATCGGAGATGCCATACAAATCATCAACGATATTAAGTTGTGTGGTAAGATAGTCTAC 883  
 Qy 384 TGAAGGCCCTGTGAAGAACTTCGAATCTGCTGCATCATTTGCCAGTTTCGTTGCAACTGC 443  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 884 CTGAAGTCCCGTAAAGTCTTTAGAGGCGCTGTCTACATTTCCGGTGACATTTGACMACTGC 943  
 Qy 444 TGG 446  
 Db 944 CGG 946

RESULT 14  
 A0873736/c

LOCUS  
 DEFINITION

A0873736 542 bp DNA linear GSS 08-NOV-1999  
 V74E1 mtn-3xHA/lacZ Insertion Library, strain AB972 Saccharomyces  
 cerevisiae genomic 5', genomic survey sequence.

ACCESSION  
 VERSION

A0873736  
 A0873736.1 GI:6285980

KEYWORDS  
 SOURCE

Saccharomyces cerevisiae (baker's yeast)  
 Saccharomyces cerevisiae  
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 Saccharomycetales; Saccharomycetaceae; Saccharomyces.

REFERENCE  
 AUTHORS

1 (bases 1 to 542)  
 Ross-Macdonald, P., Roemer, T., Coelho, P.S.R., Agarwal, S., Kumar, A.,  
 desEtages, S.A., Cheung, K.-H., Sheehan, A., Symoniat, D., Jansen, R.,  
 Umansky, L., Heidtman, M., Nelson, K., Iwasaki, H., Kanada, D., Lugo, R.,  
 Hager, K., Miller, P., Roeder, G.S. and Snyder, M.

TITLE  
 JOURNAL

Large-Scale Analysis of the Yeast Genome by Transposon Tagging and  
 Gene Disruption  
 Unpublished  
 Contact: Kumar A  
 Michael Snyder, Dept. of Mol. Cell. and Dev. Biology  
 Yale University  
 P.O. Box 208103, New Haven, CT 06520-8103, USA  
 Tel: 203 432 9949  
 Fax: 203 432 6161  
 Email: anuj.kumar@yale.edu

COMMENT

te of mtn-3xHA/lacZ insertion.  
 Seq primer: GGCCTCTTTCTTGGAGTAC  
 Class: transposon-tagged.

FEATURES  
 source

1. 542  
 /organism="Saccharomyces cerevisiae"  
 /mol\_type="genomic DNA"  
 /strain="AB972 - trp1 r(0) (S288C background)"  
 /db\_xref="taxon:4932"  
 /lab\_host="E. coli"

/clone lib="mTn-3xHA/lacZ Insertion Library, strain AB972"  
/notes="Vector: pHSS6-Sal; A yeast genomic DNA library was  
prepared in pHSS6-Sal; genomic DNA was size-fractionated  
(DNA of roughly 2-3 kb in length) prior to cloning. This  
library was subsequently mutagenized with a mTn-3xHA/lacZ  
minitransposon containing lacZ, URA3, and tet resistance."

```
BASE COUNT      152 a   136 c   112 g   142 t
ORIGIN
Query Match      10.6%; Score 117.8; DB 28; Length 542;
Best Local Similarity 63.3%; Pred. No. 1.8e-18;
Matches 198; Conservative 0; Mismatches 112; Indels 3; Gaps 1;

QY 8 TTCCAAACCACTCAAAAAGCGCTCATCTTCAAGGTGACAAAGCTGTGTTAAACAGATG 67
   |||||
DB 409 TTCCAGAAACCATGAAGCGGTGTCATTGAAATGGCGAGGCTGTAGTCAGACAGACA 350
   |||||

QY 68 TCTCAGTTCAGAAATTAAGAGGGGTACAGCCTTGGTGAAGGTGAGGCTGTGCTGGTA 127
   |||||
DB 349 TTCCAATTCTCGCATTAGAAGAAGGATTTGTTCTAATTAGACTGTGCGCGTTGCCGGTA 290
   |||||

QY 128 ACCCACTGATTCGAGCATATTCCTTATAAGATTGCTCCAGAGGTTCATATCTAGGAT 187
   |||||
DB 289 ACCATACCGATTGAAACATATTAATTTTCGAGATTGGTCTCAAGGTGCCCTCTTAGGCT 230
   |||||

QY 188 GTGACATTGCTGTACAGTTGTCAAACCTTGACCAAAATGCTAGTACTG---ACTTGAAG 244
   |||||
DB 229 GTGATGACGCGGCCAAATGTAAGTTGGGCCCAATGTTGATGCTGCACGCTTGCCA 170
   |||||

QY 245 TTGGAGATACCGGTTTCGGTTTTGTTTCAACGCTGCTCCCAACAGATCTTAATAATGGT 304
   |||||
DB 169 TTGGTGATTACATTTATGGGGTTATTCAACGCTGCTTCAGTGAGGTTCCCTCCCAACGGTG 110
   |||||

QY 305 CATTGCTGAATA 317
   |||||
DB 109 CATTGCTGAGTA 97
   |||||
```

```
RESULT 15
LOCUS      T38505
DEFINITION EST104002 S. cerevisiae strain X2180-1A Saccharomyces cerevisiae
            cDNA 5' end, mRNA sequence.
ACCESSION  T38505
VERSION     T38505.1 GI:622322
KEYWORDS    EST.
SOURCE      Saccharomyces cerevisiae (baker's yeast)
ORGANISM    Saccharomyces cerevisiae
REFERENCE   1 (bases 1 to 337)
AUTHORS     Weinstein,K.
TITLE       Saccharomyces cerevisiae cDNAs
JOURNAL     Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
COMMENT     Saccharomycetales; Saccharomycetaceae; Saccharomyces.
            Weinstock,K.
            1 (bases 1 to 337)
            Saccharomyces cerevisiae cDNAs
            Unpublished
            Contact: Weinstein,K. and Venter,J.C.
            The Institute for Genomic Research
            932 Clopper Rd, Gaithersburg, MD 20878
            Tel: 3018699056
            Fax: 3018699423
            Email: tdbinfo@tdb.tigr.org
            For clone availability please contact the TIGR Database
            (tdbinfo@tdb.tigr.org)
Seq primer: M13 Reverse.
            Location/Qualifiers
FEATURES
  source
    1..337
    /organism="Saccharomyces cerevisiae"
    /mol_type="mRNA"
    /strain="X2180-1A"
    /db_xref="taxon:4932"
  /clone_lib="S. cerevisiae strain X2180-1A"
  /notes="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
  XhoI"
```

```
BASE COUNT      83 a    75 c    77 g    97 t    5 others
ORIGIN
Query Match      10.4%; Score 115.4; DB 14; Length 337;
Best Local Similarity 63.8%; Pred. No. 6.1e-18;
Matches 203; Conservative 0; Mismatches 111; Indels 4; Gaps 2;

QY 4 TCAGTTCCAAACCACTCAAAAAGCGCTCATCTTGAAGGTGACAAAGCTGTGTTAAACA 63
   |||||
DB 2 TCAGTTCCAGAAACCATGAAGCGGTTCATTGAAATGGCAAGCGCTGTAGTCAACAG 61
   |||||

QY 64 GATGCTCTCAGTTCACAAATTAAGAGGGGTACAGCCTTGGTGAAGGTGAGGCTGTGCT 123
   |||||
DB 62 GNCATTTCCAATTCCTGAATTAGAAGAAGGATTTGTTCTAATTAAAGACTGTGCGCGTTGCC 121
   |||||

QY 124 GGTAAACCCCAACTGATTGGAAGCATATTTGCTTATAAGATTGGTCCAGAAGGTTCATTTCTA 183
   |||||
DB 122 GGTAAACCCCTACCGATTGGAAACATATTTGATTTCAAGATTGGTCTCAAGGTGCGCTCTTA 181
   |||||

QY 184 GGATGTGACATTCGCTGTGATACAGTTGTCAAACCTTGGACCAAAATCTAGTACTG---ACTTG 240
   |||||
DB 182 GGCTGTGATGTCAGCGCGGCCAAATCGTAAAGTTGGGCCCAAAATGTTNATGCTGCACGCTTT 241
   |||||

QY 241 AAGGTTGGAGATACCGGTTTCGGTTTTGTTTTCAGCGTGTCTCCCAACAGAT-CCTAAAAA 299
   |||||
DB 242 GCCATTGGTGATTACATTTATGGGTNATTCACGCTGCTTTCAGTGAGGTTCCCTCCCA 301
   |||||

QY 300 TGGTGCAATTTGCTGAATA 317
   |||||
DB 302 CGGTGNCNTTGTCTGAGTA 319
   |||||

RESULT 16
LOCUS      AZ926593
DEFINITION 476.dis09b12.s1 Saccharomyces castellii NRRL Y-12630 Saccharomyces
            castellii genomic clone 476.dis09b12.s1, genomic survey sequence.
ACCESSION  AZ926593
VERSION     AZ926593.1 GI:13497495
KEYWORDS    GSS.
SOURCE      Saccharomyces castellii
ORGANISM    Saccharomyces castellii
REFERENCE   1 (bases 1 to 540)
AUTHORS     Clifton,P.F., Hillier,L.W., Fulton,L., Graves,T., Miner,T., Gish
            ,W.R., Waterston,R.H. and Johnston,M.
TITLE       Surveying Saccharomyces genomes to identify functional elements by
            comparative DNA sequence analysis
JOURNAL     Unpublished
COMMENT     Contact: Johnston M
            Department of Genetics
            Washington University Medical School
            Box 8232, 4566 Scott Ave., St. Louis, MO 63110, USA
            Tel: 314 362 2735
            Fax: 314 362 7855
            Email: mj@genetics.wustl.edu
            Class: random plasmid subclone.
            Location/Qualifiers
FEATURES
  source
    1..540
    /organism="Saccharomyces castellii"
    /mol_type="genomic DNA"
    /strain="NRRL Y-12630 (CBS 4309)"
    /db_xref="taxon:27288"
    /clone="476.dis09b12.s1"
    /notes="Random genomic sequence"
    /notes="Random genomic sequence"

BASE COUNT      182 a    80 c    89 g    189 t

Query Match      9.7%; Score 108.4; DB 28; Length 540;
Best Local Similarity 61.9%; Pred. No. 3.9e-16;
Matches 172; Conservative 0; Mismatches 106; Indels 0; Gaps 0;
```

```
QY 828 CATTCTCTGAAGAAATCAGAAAAGATATGTTTAAATTTGATATATTTACTTTTGTGTATCGTGC 887
|||||
Db 19 CATTAAAGAGGAGACAGAGATCTAAGCTCAAGATCACCGGTACCTTGTGTATTATC 78
|||||
QY 888 ATCTGGTCAAGAAATCTATTGGGTGCAACAAGATTTCTCTGCTAGTCCAGAAATATCATGA 947
|||||
Db 79 AGTGGTTCATGACGTTCTCTTCCGTTCAATGACTTTTACCAGCTTCTCAAGAGTATAGAGA 138
|||||
QY 948 AGCCACAGTTAAATTCGTTAAAGTTTATAATTAATCCACACCTTTAAACAAGGTGATATCATCA 1007
|||||
Db 139 AGCCACTATTAAAGCTGTTTAAATTTGTTACTCCAAAGATCTTGAATGGTGAATTCATCA 198
|||||
QY 1008 TATGAATATTAAAGTTTTCAGCAACGGCTTAGATGATGTCCTCCAGCTCTCACTCAAGGTAT 1067
|||||
Db 199 TATTCATTTAAGATTTACAGATAGGTTTAAAGGATGTTCTTCAATTTGACTGATGATAT 258
|||||
QY 1068 AAAAGAGGTTAAACAAAATTTAAGTATGTTGCCA 1105
|||||
Db 259 TAAGAACATAGAAATTTCTGGGAAAATTTAGTTGCCA 296
|||||

RESULT 17
AZ926284 404 bp DNA linear GSS 01-APR-2001
LOCUS 476.dio48b07.sl Saccharomyces castellii NRRL Y-12630 Saccharomyces
DEFINITION castellii genomic clone 476.dio48b07.sl, genomic survey sequence.
ACCESSION AZ926284
VERSION AZ926284.1 GI:13497186
KEYWORDS GSS.
SOURCE Saccharomyces castellii
ORGANISM Saccharomyces castellii
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
REFERENCE 1 (bases 1 to 404)
AUTHORS Clifton, P.F., Hillier, L.W., Fulton, L., Graves, T., Miner, T., Gish,
W.R., Waterston, R.H. and Johnston, M.
TITLE Surveying Saccharomyces genomes to identify functional elements by
comparative DNA sequence analysis
JOURNAL Unpublished
COMMENT Contact: Johnston M
Department of Genetics
Washington University Medical School
Box 8232, 4566 Scott Ave., St. Louis, MO 63110, USA
Tel: 314 362 2735
Fax: 314 362 7855
Email: mj@genetics.wustl.edu
Class: random plasmid subclone.
FEATURES
source
1..404
/organism="Saccharomyces castellii"
/mol_type="genomic DNA"
/strain="NRRL Y-12630 (CBS 4309)"
/db_xref="taxon:27288"
/clone="476.dio48b07.sl"
/clone_lib="Saccharomyces castellii NRRL Y-12630"
/note="Random genomic sequence"
BASE COUNT 121 a 79 c 102 g 102 t
ORIGIN
Query Match 8.8%; Score 97.6; DB 28; Length 404;
Best Local Similarity 65.8%; Pred. No. 1.8e-13;
Matches 204; Conservative 0; Mismatches 99; Indels 7; Gaps 4;
QY 30 CATCATGAAGGTGACAAAGCTGTTGTTAAACAGATGTTCTAGTTCAGTTCAGAAATTAAGGA 89
|||||
Db 1 CGTCATGAAGGCGATAAATCAATACTCAAGACAGGTGTGCCCTTACCACCTTTGGAAG 60
|||||
QY 90 GGCTACAGCCCTCGTGAAGGTTTGAGCTGTGCTGTTAACCCTCACTGATTGGAGCATAT 149
|||||
Db 61 TGGTTCGTTCTCGTGAAGACGAAAGCAGTGTCTGGGAACCC-ACAGATTGGAACATAT 119
|||||
QY 150 TGCTTTAAG--ATTGGTCAGAAAGGTTTCAATTTCTAG-GATGTGACATTTGCTGGTACAGT 206
|||||
```

```
Db 120 AGATTTCAGGATATTGGACCACCAAGTTCCATTCTTGAGTTTGACATCGCTGGTCAAT 179
|||||
QY 207 TGTCAAACTTGGACCACCAAT--GCTAGTACTGACTTGAAGTTGGAGATACCGGTTTCGG 263
|||||
Db 180 TGTGAAATTTGGGACCCGCGATGTAGACCCCAATGAATTTCCATATTTGGTGATATGTTTGG 239
|||||
QY 264 TTTTGTTCACGGTGTCTTCCAAACAGATCCTTAAATGTTGCAATTTGCTGGAATATGCCAG 323
|||||
Db 240 GTTTGTACATGGTGTCTTCCATTAAACGTCCCAAAATGGTGCCTTTGCAGAGTATGCTGT 299
|||||
QY 324 GGTTTATCCA 333
|||||
Db 300 CATGGATCCA 309
|||||

RESULT 18
AQ501976/c 353 bp DNA linear GSS 29-APR-1999
LOCUS V11A3 mTn-3xHA/lacZ Insertion library Saccharomyces cerevisiae
DEFINITION genomic 5', genomic survey sequence.
ACCESSION AQ501976
VERSION AQ501976.1 GI:4707626
KEYWORDS GSS.
SOURCE Saccharomyces cerevisiae (baker's yeast)
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
REFERENCE 1 (bases 1 to 353)
AUTHORS Ross-Macdonald, P., Roemer, T., Coelho, P.S.R., Agarwal, S., Kumar, A.,
desBages, S.A., Cheung, K.-H., Sheehan, A., Symoniat, D., Jansen, R.,
Umansky, L., Heidman, M., Nelson, K., Iwasaki, H., Kanada, D., Lugo, R.,
Hager, K., Miller, P., Roeder, G.S. and Snyder, M.
TITLE Large-Scale Analysis of the Yeast Genome by Transposon Tagging and
Gene Disruption
JOURNAL Unpublished
COMMENT Contact: Kumar A
Michael Snyder, Dept. of Mol. Cell. and Dev. Biology
Yale University
P.O. Box 208103, New Haven, CT 06520-8103, USA
Tel: 203 432 9949
Fax: 203 432 6161
Email: anuj.kumar@yale.edu
te of mTn-3xHA/lacZ insertion.
Seq primer: GGCCTCTCTTCTTTGGAAGTAC
Class: transposon-tagged.
FEATURES
source
1..353
/organism="Saccharomyces cerevisiae"
/mol_type="genomic DNA"
/db_xref="taxon:4932"
/lab_host="E. coli"
/clone_lib="mTn-3xHA/lacZ Insertion Library"
/note="Vector: pHS86-Sal; A yeast genomic DNA library
(lacking mitochondrial DNA) was prepared in pHS86-Sal;
genomic DNA was size-fractionated (DNA of roughly 2-3 kb
in length) prior to cloning. This library was
subsequently mutagenized with a mTn-3xHA/lacZ
minitransposon containing lacZ, URA3, and tet resistance."
BASE COUNT 95 a 95 c 76 g 84 t
ORIGIN
Query Match 8.7%; Score 96.4; DB 28; Length 353;
Best Local Similarity 64.4%; Pred. No. 3.4e-13;
Matches 161; Conservative 0; Mismatches 86; Indels 3; Gaps 1;
QY 71 CAGTTCAGAAATTAAGAGCGGTACAGCCCTTGTTGAAGTTGAGGCTGTTGCTGGTAACC 130
|||||
Db 352 CAATTCCTGAATTAGAGAGGATTTGTTCTAATTAAGACTGTGCGCGTTGCGGTAAAC 293
|||||
QY 131 CAACTGATTGGAAGCATATTTGCTTTAAGATTGGTCCAGAGGTTCAATTTAGGATGTG 190
|||||
Db 292 CTACCGATTGGAACAATATTTGATTTCAAGATTGGTCTCTCAAGGTGCCCTCTTAGGCTGTG 233
|||||
```



Qy 191 ACATTGCTGGTACAGTTGTCAAACTTGGACCAATGCTAGTACTG---ACTTGAAGGTTG 247  
 Db 232 ATGCAGCGCGCAATCGTAAAGTTGGGCCCAATGTTGATGTTGCACGCTTTGCCATTG 173  
 Qy 248 GAGTACCGGTTTCGGTTTGTTCACGGTCTTCCAAACAGATCCTAAAAATGGTGCAT 307  
 Db 172 GTGATACATTATATGGGTTATTTCACGGTGTTCAGTGAGGTTCCCTCAACCGTGCCCT 113  
 Qy 308 TTGCTGAATA 317  
 Db 112 TTGCTGAGTA 103

RESULT 19  
 LOCUS AQ876344/c 549 bp DNA linear GSS 08-NOV-1999  
 DEFINITION V98D9 mTn-3xHA/lacZ Insertion Library, strain Y2278 Saccharomyces cerevisiae genomic 5', genomic survey sequence.  
 ACCESSION AQ876344  
 VERSION AQ876344.1 GI:6288588  
 KEYWORDS GSS.  
 SOURCE Saccharomyces cerevisiae (baker's yeast)  
 ORGANISM Saccharomyces cerevisiae  
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 REFERENCE 1 (bases 1 to 549)  
 AUTHORS Ross-Macdonald, P., Roemer, T., Coelho, P. S. R., Agarwal, S., Kumar, A., deBates, S. A., Cheung, K. H., Sheehan, A., Symonistis, D., Jansen, R., Unanysky, L., Heidman, M., Nelson, K., Iwasaki, H., Kanada, D., Lugo, R., Hager, K., Miller, P., Roeder, G. S. and Snyder, M.  
 TITLE Large-Scale Analysis of the Yeast Genome by Transposon Tagging and Gene Disruption  
 JOURNAL Unpublished  
 COMMENT Contact: Kumar A  
 Michael Snyder, Dept. of Mol. Cell. and Dev. Biology  
 Yale University  
 P.O. Box 208103, New Haven, CT 06520-8103, USA  
 Tel: 203 432 9949  
 Fax: 203 432 6161  
 Email: anuj.kumar@yale.edu  
 te of mTn-3xHA/lacZ insertion.  
 Seq primer: GGCCTCTCTCTTTGGAAGTAC  
 Class: transposon-tagged.

FEATURES  
 Source  
 1. 549  
 /organism="Saccharomyces cerevisiae"  
 /mol\_type="genomic DNA"  
 /strain="Y2278 - S288C background, cir(0) rho(0)"  
 /db\_xref="taxon:4932"  
 /lab\_host="E. coli"  
 /clone\_lib="mTn-3xHA/lacZ Insertion Library, strain Y2278"  
 /note="Vector: pHS6-Sal; A yeast genomic DNA library without 2 micron or mitochondrial DNA was prepared in pHS6-Sal; genomic DNA was size-fractionated (DNA of roughly 2-3 kb in length) prior to cloning. This library was subsequently mutagenized with a mTn-3xHA/lacZ minitransposon containing lacZ, URA3, and tet resistance."

BASE COUNT 141 a 120 c 109 g 179 t  
 ORIGIN  
 Query Match 8.1%; Score 90.2; DB 28; Length 549;  
 Best Local Similarity 57.7%; Pred. No. 1.4e-11;  
 Matches 161; Conservative 0; Mismatches 118; Indels 0; Gaps 0;  
 Qy 835 GAAGAAATCAGAAAGATAATGTTAAATTTGATATTTCTTTGTTGTCGTCATCGT 894  
 Db 544 GAGGAGACAGGCGCAACGTCAGTATTGAAGAACCTTCTAAATTTGATAGGAGT 485  
 Qy 895 CAAGAAATTCATTGGGTGCAACAGATTTCTGCTAGTCAGAAATATCATGAAGCCACA 954  
 Db 484 AACACGCTCCCATTTTGGAACTTTTACTTTTGCAGCAGACCCCTGAATCAAGGAGCGGCC 425

Qy 955 GTTAAATTCGTTAAAGTTTATAAATCCACACCTTTAAACAACGGTGATATCCATCATATGAAT 1014  
 Db 424 ATAAATTTATTAAAGTTTCATCAATCCAAAATCAATGATGGTGAATCCACCATCCCA 365  
 Qy 1015 ATTAAAGTTTTCAGCAACGGCTTAGATGATGATCCCGAGCTCTCACTGAAGGTATAAAAGAA 1074  
 Db 364 GTGATAGTTTACAAAGAACGGGTTAGATGATATATCCACAGTTACTTGTATGATATTAAGCAC 305  
 Qy 1075 GGTAAAAACAAAATGTTAAGTATGTTGCCAGGTTATAA 1113  
 Db 304 GGGAGGAATTCGGCGAAAAGTTGGTTGGCTTGGAA 266

RESULT 20  
 LOCUS AZ930225 522 bp DNA linear GSS 01-APR-2001  
 DEFINITION 474.dhz51b06.61 Saccharomyces unisporus NRRL Y-1556 Saccharomyces unisporus genomic clone 474.dhz51b06.61, genomic survey sequence.  
 ACCESSION AZ930225  
 VERSION AZ930225.1 GI:13501134  
 KEYWORDS GSS.  
 SOURCE Saccharomyces unisporus  
 ORGANISM Saccharomyces unisporus  
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 REFERENCE 1 (bases 1 to 522)  
 AUTHORS Clifton, P. F., Hillier, L. W., Fulton, L., Graves, T., Miner, T., Gish, W. R., Waterston, R. H. and Johnston, M.  
 TITLE Surveying Saccharomyces genomes to identify functional elements by comparative DNA sequence analysis  
 JOURNAL Unpublished  
 COMMENT Contact: Johnston M  
 Department of Genetics  
 Washington University Medical School  
 Box 8232, 4566 Scott Ave., St. Louis, MO 63110, USA  
 Tel: 314 362 2735  
 Fax: 314 362 7855  
 Email: mj@genetics.wustl.edu  
 Class: random plasmid subclone.

FEATURES  
 Source  
 1. 522  
 /organism="Saccharomyces unisporus"  
 /mol\_type="genomic DNA"  
 /strain="NRRL Y-1556 (CBS 398)"  
 /db\_xref="taxon:27294"  
 /clone="474.dhz51b06.61"  
 /clone\_lib="Saccharomyces unisporus NRRL Y-1556"  
 /notes="Random genomic sequence"

BASE COUNT 162 a 77 c 90 g 193 t  
 ORIGIN  
 Query Match 7.2%; Score 79.6; DB 28; Length 522;  
 Best Local Similarity 66.1%; Pred. No. 5.9e-09;  
 Matches 115; Conservative 0; Mismatches 59; Indels 0; Gaps 0;  
 Qy 1 ATGTCAGTTCCCAACCACTCAAAAGCGTCATCATTTGAAGGTGACAAAGCTGTTGTTAAA 60  
 Db 349 ATGTCCTTACTACTACCATGAGGCGCTGTTGTCGAAGATGACCATGTTGTTGTCAG 408  
 Qy 61 ACAGATGTCCTCAGTTCAGAAATTTAAAGGAGGGGTACAGCCCTTGGTGAAGGTTGAGGCTGTT 120  
 Db 409 GATAACGTCCTCATTTACCAGAAATTTAGAAGATGGGTTTTTTATTAGTGAACCAAGGCTGTC 468  
 Qy 121 GCTGGTAAACCAACTGATTGGAAGCATATTTGCTTATTAAGATTGGTCAGAAAGT 174  
 Db 469 GCAGGTAATCAACTGATTGGAACATATTTGCTTTTAAAGATGTCCTCCACAAGAT 522

RESULT 21  
 LOCUS BZ293919 574 bp DNA linear GSS 31-OCT-2002  
 DEFINITION CG0601.r1 Candida glabrata Random Genomic Library Candida glabrata

genomic clone CG0601, genomic survey sequence.

ACCESSION BZ293919  
VERSION BZ293919.1 GI:24434218  
KEYWORDS GSS.  
SOURCE Candida glabrata  
ORGANISM Candida glabrata  
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
Saccharomycetales; mitosporic Saccharomycetales; Candida.  
1 (bases 1 to 574)  
REFERENCE Wong, S., Fares, M.A., Zimmermann, W., Butler, G. and Wolfe, K.H.  
AUTHORS Evidence from comparative genomics for a complete sexual cycle in  
TITLE the 'asexual' pathogenic yeast *Candida glabrata*  
JOURNAL Genome Biol. 4 (2), R10 (2003)  
MEDLINE 22508158  
PUBMED 12620120  
COMMENT Contact: Wong S  
Department of Genetics, Smurfit Institute  
Trinity College Dublin  
Dublin 2, Ireland  
Tel: 353 1 6082319  
Fax: 353 1 6798558  
Email: swong@tcd.ie  
Class: plasmid ends.

FEATURES  
source  
1. 574  
/organism="Candida glabrata"  
/mol\_type="genomic DNA"  
/strain="CBS 138"  
/db\_xref="taxon:5478"  
/clones="CG0601"  
/clone\_lib="Candida glabrata Random Genomic Library"  
BASE COUNT 183 a 104 c 106 g 181 t  
ORIGIN  
Query Match 6.6%; Score 73.2; DB 29; Length 574;  
Best Local Similarity 69.7%; Pred. No. 2.4e-07;  
Matches 99; Conservative 0; Mismatches 43; Indels 0; Gaps 0;  
QY 60 AACAGATGTCAGTTCAGATTAAGAGGGGTACAGCCTTGGTGAAGTTGAGGCTGT 119  
|||  
DB 429 AAGGAAAGTTCCTGTCAGAAATTCGCGTGGTGTCTAGTTAAGAAATAGAGCTGT 488  
|||  
QY 120 TGCTGGTAACCAACTGATGGAAGCATATGCTTATAAGATGTCACAGAGTTTCAAT 179  
|||  
DB 489 CGCTGGTAACCAACTGGAAGCATACATGGAAGATGACACACAGTTCCAT 548  
|||  
QY 180 TCTAGGATGACATTCCTGT 201  
|||  
DB 549 CGCGGTTGTGACATTCGCGT 570  
|||

RESULT 22  
AQ501614/c  
LOCUS  
DEFINITION V17F2 mTn-3xHA/lacZ Insertion Library Saccharomycetes cerevisiae  
genomic 5', genomic survey sequence.  
ACCESSION AQ501614  
VERSION AQ501614.1 GI:4707264  
KEYWORDS GSS.  
SOURCE Saccharomycetes cerevisiae (baker's yeast)  
ORGANISM Saccharomycetes cerevisiae  
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
Saccharomycetales; Saccharomycetales; Saccharomycetes.  
1 (bases 1 to 454)  
REFERENCE Ross-Macdonald, P., Roemer, T., Coelho, P.S.R., Agarwal, S., Kumar, A.,  
AUTHORS desEtages, S.A., Cheung, K.-H., Sheehan, A., Symonitis, D., Jansen, R.,  
Unanue, L., Heidtman, M., Nelson, K., Iwasaki, H., Kanada, D., Lugo, R.,  
Hager, K., Miller, P., Roeder, G.S. and Snyder, M.  
TITLE Large-Scale Analysis of the Yeast Genome by Transposon Tagging and  
Gene Disruption  
JOURNAL Unpublished  
COMMENT Contact: Kumar A  
Michael Snyder, Dept. of Mol. Cell. and Dev. Biology

Yale University  
P.O. Box 208103, New Haven, CT 06520-8103, USA  
Tel: 203 432 9949  
Fax: 203 432 6161  
Email: anuj.kumar@yale.edu  
te of mTn-3xHA/lacZ insertion.  
Seq primer: GGCTCTCTCTCTTGAAGTAC  
Class: transposon-tagged  
Location/Qualifiers  
1. 454  
/organism="Saccharomycetes cerevisiae"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:4932"  
/lab\_host="E. coli"  
/clone\_lib="mTn-3xHA/lacZ Insertion Library"  
/notes="Vector: pHS86-Sal; A yeast genomic DNA library  
(lacking mitochondrial DNA) was prepared in pHS86-Sal;  
genomic DNA was size-fractionated (DNA of roughly 2-3 kb  
in length) prior to cloning. This library was  
subsequently mutagenized with a mTn-3xHA/lacZ  
minitransposon containing lacZ, URA3, and tet resistance."  
BASE COUNT 123 a 114 c 71 g 145 t  
ORIGIN

Query Match 5.5%; Score 61.2; DB 28; Length 454;  
Best Local Similarity 65.2%; Pred. No. 0.00022;  
Matches 90; Conservative 0; Mismatches 48; Indels 0; Gaps 0;  
QY 8 TTCCAACTCTCAAAAGCGTCATCATTTGAAGGTGACAAAGCTGTGTAAACAGATG 67  
|||  
DB 177 TTCCAGAAACCCAGAAACCGTTTTCATGAAATGCAAGGCTGTAGTCAACAGACA 118  
|||  
QY 68 TCTCAGTTCCAGAAATTAAGAGGGGTACAGCCTTGGTGAAGTTGAGGCTGTCTGGTA 127  
|||  
DB 117 TTCCAATTCCTGAATTAGAAGGATTTGTTCTAATTAAGACTGTCGCGTTGCCGTA 58  
|||  
QY 128 ACCCAACTGATTCGAAGC 145  
|||  
DB 57 ACCCTACGATTTGGAAC 40  
|||

RESULT 23  
CNS07BB3  
LOCUS  
DEFINITION T3 end of clone BC0AA010F11 of library BC0AA from strain CBS 767 of  
Debaryomyces hansenii, genomic survey sequence.  
ACCESSION AL437653  
VERSION AL437653.1 GI:12221066  
KEYWORDS GSS.  
SOURCE Debaryomyces hansenii (anamorph: Candida famata)  
ORGANISM Debaryomyces hansenii  
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
Saccharomycetales; Saccharomycetales; Debaryomyces.  
1 (bases 1 to 1146)  
REFERENCE Souciet, J.L., Aigle, M., Artiguenave, F., Blandin, G.,  
AUTHORS Bolotin-Fukuhara, M., Bon, E., Brothier, P., Casaregola, S.,  
de-Montigny, J., Dujon, B., Durand, P., Lepingle, A., Lorente, B.,  
Malpertuy, A., Neuveglise, C., Ozier-Kalogeropoulos, O., Potier, S.,  
Saurin, W., Tekala, F., Toffano-Nioche, C., Wesolowski-Louvel, M.,  
Wincker, P. and Weissenbach, J.  
TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of  
yeast species for molecular evolution studies  
FEBS Lett. 487 (1), 3-12 (2000)  
JOURNAL 20584711  
MEDLINE 11152876  
PUBMED 11152876  
REFERENCE 2 (bases 1 to 1146)  
AUTHORS Lepingle, A., Casaregola, S., Neuveglise, C., Bon, E., Nguyen, H.,  
Artiguenave, F., Wincker, P. and Gaillardin, C.  
TITLE Genomic exploration of the hemiascomycetous yeasts: 14.  
Debaryomyces hansenii var. hansenii  
FEBS Lett. 487 (1), 82-86 (2000)  
JOURNAL 20584724  
MEDLINE 11152889  
PUBMED 11152889

```

REFERENCE 3 (bases 1 to 1146)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex - FRANCE (E-mail :
seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT This GSS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.
FEATURES
    source
        1..1146
            /organism="Debaryomyces hansenii"
            /mol_type="genomic DNA"
            /strain="CBS 767"
            /variety="hansenii"
            /db_xref="taxon:4959"
            /clone="BC0AA010F11"
            /clone_lib="BC0AA"
            /note="end : T3"
        <2..>550
            /note="similar to Saccharomyces cerevisiae ORF YNL134C [
            similarity to C.carbonum toxD gene ]"
            /evidence=not_experimental
        <14..>550
            /note="similar to Saccharomyces cerevisiae ORF YCR102C [
            similarity to C.carbonum toxD gene ]
            similar to Saccharomyces cerevisiae ORF YLR460C [
            similarity to C.carbonum toxD protein ]"
            /evidence=not_experimental
        <777..>1145
            /note="similar to Saccharomyces cerevisiae ORF YBR034C [
            HMT1 ; hRNP methyltransferase ]"
            /evidence=not_experimental
    BASE COUNT 363 a 185 c 212 g 383 t 3 others
    ORIGIN
        Query Match 5.0%; Score 56.2; DB 29; Length 1146;
        Best Local Similarity 55.2%; Pred. No. 0.0051;
        Matches 153; Conservative 0; Mismatches 118; Indels 6; Gaps 2;
    Qy 558 AATCAAAGTTCGCAACATATCAATGCTTATCTAAGATTGTAATCTGTTCTTAAAAA 617
        |||||
        Db 13 AATCAAAGTCGTAAGCTTGCATATGTTT---TAAAGTCATTACAACCGCATCCTCAA 69
    Qy 618 GCATGAAAGCTTTTAAAGCTTATGTCGTGATGATGCTTTGACTATCATGATGAGG 677
        |||||
        Db 70 GCATCAGATTTCTGAAATCGTTAGTGTGATGAAAGTATTTGATTACCATGACTCAA 129
    Qy 678 CGTTATTGAGCAGAT---CAATCGAAGTATCCAAACCTGCAACATGTTATTGACGCTGT 734
        |||||
        Db 130 TGTCATTGAGCAGATGAAGAAATCTGGTGGTCTTAACATCACTACGCATTGGATCTGT 189
    Qy 735 GGGAAAGCGAAGATAGTATCCCGAGGCGCTATAAAGTCACAGCAGATGCTTACCTGCCAC 794
        |||||
        Db 190 TTCCAACGAACAACTTTCCAATCGGTACGATGTCACGCGCTGATCACTAAGGATGTTGC 249
    Qy 795 ATTATTAGAAGTGGTTCGAATGACCATGGAAGCATT 831
        |||||
        Db 250 TATTGATAAATTGCTCTTTTAACTCTCTGACAAAAT 286
RESULT 24
LOCUS BQ143479 556 bp mRNA linear EST 24-APR-2002
DEFINITION fmlc.pk005.p23 Metarhizium anisopliae sf. acridum ARSEF 324
Metarhizium anisopliae var. acridum cDNA, mRNA sequence.
REFERENCE 1 (bases 1 to 515)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex - FRANCE (E-mail :
seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT This GSS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.
FEATURES
    source
        1..556
            /organism="Metarhizium anisopliae var. acridum"
            /mol_type="mRNA"
            /strain="ARSEF 324"
            /db_xref="taxon:92637"
            /clone_lib="Metarhizium anisopliae sf. acridum ARSEF 324"
            /notes="Vector: UniZap; Metarhizium anisopliae sf. acridum
            was grown on insect cuticle and chitin for 24 hours. A
            cDNA library was constructed in the unidirectional Lambda
            vector UniZap."
        BASE COUNT 147 a 134 c 145 g 123 t 7 others
        ORIGIN
            Query Match 4.7%; Score 52.2; DB 13; Length 556;
            Best Local Similarity 53.4%; Pred. No. 0.041;
            Matches 133; Conservative 0; Mismatches 113; Indels 3; Gaps 1;
        Qy 68 TCTCAGTTCAGAAATTTAAAGGGGTACAGCCCTTGGTGAAGGTTGAGGCTGTTGCTGGTA 127
            |||||
            Db 88 TCCCGTGCACATTTACCTGATGATCTACTTATTCGCTGTTGTCAGTGGCGTAA 147
        Qy 128 ACCCACTGATTGGAAGCATATTTGCTTATAAGATTGGTCCAGAGGTTTCAATTTCTAGGAT 187
            |||||
            Db 148 ATCCCACTGACTGACACAAACGCTTG---ATGCACCGGTGACGATGCGACGCTGCTAGGAT 204
        Qy 188 GTGACATTTGCTGTACAGTTGTCAAACTTGGACCAATGCTAGTACTGACTTGAAGGTTG 247
            |||||
            Db 205 GTGACTGGGCGAGAACAGTAGAAGAAATTTGGGCCCAAGGTGAAAAGGCCATTCCAAATAG 264
        Qy 248 GAGATACCGGTTTTCGGTGTGTTTTCACGGTGTCTTCCCAACAGATCCTTAAAAATGGTGCAT 307
            |||||
            Db 265 GTGAAAGGTTTTCGGGCTTTGTCGAGGATTAATGATGCGAGACACACACGCGGAGCTT 324
        Qy 308 TTGCTGAAT 316
            |||||
            Db 325 TCGCAAAAT 333
RESULT 25
LOCUS BQ143479 515 bp mRNA linear EST 08-APR-2002
DEFINITION BB160022A20C02 5 Bee Brain Normalized Library, BB16 Apis mellifera
cDNA clone BB160022A20C02 5', mRNA sequence.
REFERENCE 1 (bases 1 to 515)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex - FRANCE (E-mail :
seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT This GSS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.
FEATURES
    source
        1..515
            /organism="Apis mellifera (honeybee)"
            /mol_type="mRNA"
            /strain="BB160022A20C02 5'"
            /db_xref="taxon:6068"
            /clone_lib="Bee Brain Normalized Library, BB16 Apis mellifera"
            /notes="Vector: UniZap; Metarhizium anisopliae sf. acridum
            was grown on insect cuticle and chitin for 24 hours. A
            cDNA library was constructed in the unidirectional Lambda
            vector UniZap."
        BASE COUNT 147 a 134 c 145 g 123 t 7 others
        ORIGIN
            Query Match 4.7%; Score 52.2; DB 13; Length 556;
            Best Local Similarity 53.4%; Pred. No. 0.041;
            Matches 133; Conservative 0; Mismatches 113; Indels 3; Gaps 1;
        Qy 68 TCTCAGTTCAGAAATTTAAAGGGGTACAGCCCTTGGTGAAGGTTGAGGCTGTTGCTGGTA 127
            |||||
            Db 88 TCCCGTGCACATTTACCTGATGATCTACTTATTCGCTGTTGTCAGTGGCGTAA 147
        Qy 128 ACCCACTGATTGGAAGCATATTTGCTTATAAGATTGGTCCAGAGGTTTCAATTTCTAGGAT 187
            |||||
            Db 148 ATCCCACTGACTGACACAAACGCTTG---ATGCACCGGTGACGATGCGACGCTGCTAGGAT 204
        Qy 188 GTGACATTTGCTGTACAGTTGTCAAACTTGGACCAATGCTAGTACTGACTTGAAGGTTG 247
            |||||
            Db 205 GTGACTGGGCGAGAACAGTAGAAGAAATTTGGGCCCAAGGTGAAAAGGCCATTCCAAATAG 264
        Qy 248 GAGATACCGGTTTTCGGTGTGTTTTCACGGTGTCTTCCCAACAGATCCTTAAAAATGGTGCAT 307
            |||||
            Db 265 GTGAAAGGTTTTCGGGCTTTGTCGAGGATTAATGATGCGAGACACACACGCGGAGCTT 324
        Qy 308 TTGCTGAAT 316
            |||||
            Db 325 TCGCAAAAT 333

```

**AUTHORS** Whitfield, C.W., Band, M.R., Bonaldo, M.F., Kumar, C.G., Liu, L.,  
 Pardinas, J., Robertson, H.M., Soares, B. and Robinson, G.E.  
**TITLE** Annotated expressed sequence tags and cDNA microarrays for studies  
 of brain and behavior in the honey bee  
**JOURNAL** Genome Res. 12 (4), 555-566 (2002)  
**MEDLINE** 21929762  
**PUBMED** 11932240  
**COMMENT** Contact: Gene E. Robinson  
 Department of Entomology  
 University of Illinois  
 505 S. Goodwin Ave., Urbana, IL 61801, USA  
 Tel: 217 265 0309  
 Fax: 217 244 3499  
 Email: generobi@life.uiuc.edu  
 This research was funded by the University of Illinois Critical  
 Research Initiatives Fund and a Burroughs-Wellcome Trust Innovation  
 Award in Functional Genomics to G.E. Robinson and an NSF  
 Postdoctoral Fellowship in Bioinformatics to C.W. Whitfield.  
**PCR PRIMERS**  
 FORWARD: TAATACGACTCACTATAGG  
 BACKWARD: ATTAACCTCACTAAG  
 Plate: BL60022A20 row: C column: 02  
 Seq primer: AGCGATAACAAATTCACACAGGA  
 High quality sequence stop: 515.

# **FEATURES** source

1. 515  
 /organism="Apis mellifera"  
 /mol\_type="mRNA"  
 /strain="mixed strains of European bees, predominantly  
 A.m. ligustica"  
 /db\_xref="taxon:7460"  
 /clone="BB160022A20C02"  
 /sex="female"  
 /tissue\_type="brain"  
 /dev\_stage="adult worker honey bee"  
 /lab\_host="DH10B"  
 /clone\_lib="Bee Brain Normalized Library, BB16"  
 /note="Organ: brain; Vector: pT73-Pac; Site 1: EcoRI;  
 Site 2: NotI; The BB16 library was contributed by the  
 Soares laboratory and it was constructed and normalized  
 as described by Bonaldo, M.F., Lennon, G. and Soares,  
 M.B. (1996), Genome Research 6(9): 791-806. RNA was  
 prepared from dissected brains of adult worker bees of  
 various ages and various behavioral groups."  
 198 a 74 c 89 g 153 t 1 others

# **BASE COUNT** ORIGIN

Query Match 4.6%; Score 51.6; DB 12; Length 515;  
 Best Local Similarity 52.8%; Pred. No. 0.056;  
 Matches 134; Conservative 0; Mismatches 119; Indels 1; Gaps 1;  
 Qy 766 AAGTCAAGCAGATAGTCTACCTGCCACATATTAGAGTGGTTCCAAATGACCATTTGAA 825  
 Db 129 AAATTCATCTTTGATTTGCTCATACATCATTTATATCAAGGGCTCAAGTACTCTTTAA 188  
 Qy 826 AGCATTCCTGAAGAAATCAGAAAGATAAGTTAAATTTGATATTTACTTTGTTGTTATCGT 885  
 Db 189 GCAATTTCTTAAGAAATAGGACAAAGAAACATTACTGTTCAAAGACATGGCGATTAAAT 248  
 Qy 886 GCATCTGGTCAAGAAATTTCTATTTGGGTGCAACAGATTTCTGCTAGTCCAGAAATATCAT 945  
 Db 249 GAACGTCATTATGAGGTTTGACTGGTTTAAATAAAGCTGAAACTGCTGCTAATATAGT 308  
 Qy 946 GAAGCCACAGTT-AAATTCGTTAAGTTTATAAATCCACCTTAAACACGGTGATATCCA 1004  
 Db 309 GAAGAACAAAGTTCAAATTTGGAGAAGATCTTTTGATACACCTCTCTCCACCTATGGAACCA 368  
 Qy 1005 TCATATCAATATTA 1018  
 Db 369 GATCATTAATATTA 382

RESULT 26

BP099113  
 LOCUS  
 DEFINITION BP099113 Chaetomium cDNA library Chaetomium globosum cDNA clone  
 high chaeb.Contig224, mRNA sequence.  
 ACCESSION BP099113  
 VERSION BP099113.1 GI:27438553  
 KEYWORDS EST.  
 SOURCE Chaetomium globosum  
 ORGANISM Chaetomium globosum  
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 Sordariomycetidae; Sordariales; Chaetomiaceae; Chaetomium.  
 REFERENCE 1 (bases 1 to 595)  
 AUTHORS Yang, Q. and Jin, H.  
 TITLE Chaetomium mycelium EST  
 JOURNAL Unpublished  
 COMMENT Contact: Qian Yang  
 Department of Biotechnology  
 Harbin Institute of Technology  
 West dazhi 92, Harbin, Heilongjiang 150001, China  
 Tel: 86-451-563-5236  
 Fax: 86-451-641-2952  
 Email: kimhongseng@x263.net.

# **FEATURES** source

1. 595  
 /organism="Chaetomium globosum"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:38033"  
 /clone="high chaeb.Contig224"  
 /dev\_stage="mycelium"  
 /clone\_lib="Chaetomium cDNA library"  
 125 a 214 c 163 g

# **BASE COUNT** ORIGIN

Query Match 4.6%; Score 51.6; DB 12; Length 595;  
 Best Local Similarity 47.5%; Pred. No. 0.058;  
 Matches 153; Conservative 0; Mismatches 169; Indels 0; Gaps 0;  
 Qy 1 ATGTCAAGTTCACCACTCAAAGCCGTCATCATTTGAAGTGACAAAGCTGTTGTATAA 60  
 Db 57 ATGCGCCCGCCATCAGAGATCAATGCGGTGTCATCACCAGCCCGAAGCGCTGAAGTC 116  
 Qy 61 ACAGATGTCTCAGTTCCAGAAATTAAGGAGGGGTACAGCCTTGTGAAGGTTGAGGCTGTT 120  
 Db 117 AAGACGGTACCCCTCCAGAGCTCCCGGATGACTACATCTCTCGCCACCGGCGG 176  
 Qy 121 GTGTGTAACCCCACTGATTGGAAGCATATGCTTATAAGATTGGTCCAGAAGTTCAATT 180  
 Db 177 GCTCTCAACCCGACAGACTGGAAGCAGCTTCACGTGGCGGTGAGCACCCTAGGACCCGT 236  
 Qy 181 CTAGGATGTGACATTCCTGTTACAGTTGTCAAACTTGGACCAATGCTAGTACTGACTTG 240  
 Db 237 GTAGGCTGGGACTACCGCGCATCTCGGGAATTCGCGAGAGTGGGCCCCAAGTGAACAGCCCTTC 296  
 Qy 241 AAGGTTGGAGATACCGGTTTCGGTTTTCGTTTTCACGGTGTTCCTCCAAACAGATCTCTAAAAAT 300  
 Db 297 AAAAGGGTGAACGTTATCTCGGAATTTGCCAGGAGCCAACTCGCTGCGGCTGTATGCG 356  
 Qy 301 GGTGCAATTTGCTGAATATGCCA 322  
 Db 357 GGTGGTTTTTCGCGACTACATCA 378

# **RESULT 27** BI505803 LOCUS

DEFINITION BI505803 619 bp mRNA linear EST 08-APR-2002  
 BBI70024A10H01.5 Bee Brain Normalized/Subtracted Library, BBI7 Apis  
 mellifera cDNA clone BBI70024A10H01.5, mRNA sequence.  
 ACCESSION BI505803  
 VERSION BI505803.1 GI:15356177  
 KEYWORDS EST.  
 SOURCE Apis mellifera (honeybee)  
 ORGANISM Apis mellifera  
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Apoidea; Apidae; Apis.

1 (bases 1 to 619)

Whitfield, C.W., Band, M.R., Bonaldo, M.F., Kumar, C.G., Liu, L., Pardinas, J., Robertson, H.M., Soares, B. and Robinson, G.E.

Annotated expressed sequence tags and cDNA microarrays for studies of brain and behavior in the honey bee

Genome Res. 12 (4), 555-566 (2002).

21929762

11932240

Contact: Gene E. Robinson  
Department of Entomology  
University of Illinois  
505 S. Goodwin Ave., Urbana, IL 61801, USA  
Tel: 217 265 0309  
Fax: 217 244 3499  
Email: generobi@life.uiuc.edu

This research was funded by the University of Illinois Critical Research Initiatives Fund and a Burroughs-Wellcome Trust Innovation Award in Functional Genomics to G.E. Robinson and an NSF Postdoctoral Fellowship in Bioinformatics to C.W. Whitfield.

PCR Primers

FORWARD: TAATACGACTCACTATAGG  
BACKWARD: ATTAACCCCTCACTAAG

Plate: BB170024A10 row: H column: 01

Seq primer: AGCGGATAACAATTCACACAGGA

High quality sequence stop: 619.

Location/Qualifiers

1..619

/organism="Apis mellifera"

/mol\_type="mRNA"

/strain="mixed strains of European bees, predominantly A.m. ligustica"

/db\_xref="taxon:7460"

/clone="BB170024A10H01"

/sex="female"

/tissue\_type="brain"

/dev\_stage="adult worker honey bee"

/lab\_host="DH10B"

/clone\_lib="Bee Brain Normalized/Subtracted Library, BB17"

/note="Organ: Brain; Vector: pT73-Pac; Site: 1: EcoRI; Site 2: NotI; This BB17 cDNA library was generated by subtraction of the BB16 library with 4000 previously sequenced clones. The BB16 library was contributed by the Soares laboratory and it was constructed and normalized as described by Bonaldo, M.F., Lennon, G. and Soares, M.B. (1996). Genome Research 6(9): 791-806. RNA was prepared from dissected brains of adult worker bees of various ages and various behavioral groups."

BASE COUNT 245 a 85 c 111 g 178 t

ORIGIN

Query Match 4.6%; Score 51.6; DB 12; Length 619;  
Best Local Similarity 52.8%; Pred. No. 0.059;  
Matches 134; Conservative 0; Mismatches 119; Indels 1; Gaps 1;

QY 766 AAGTTCACACGACATGTCTACCTGCCACATTTAGAGTGTTCACATGACATTGAA 825  
|||||

Db 158 AATTACATTTGATATGCTCATACATCAATTAATCAAGGGCTCAAGATCTCTAAA 217  
|||||

QY 826 AGCATTCCTGAAGAAATCAGAAAGATAATGTTAAATTTGATATTAATCTTTGTTGATCGT 885  
|||||

Db 218 GCAATTCCTTAAGAAATAGCAAGAAACATTAATCTGTTCAAAGACATCGCATTAAT 277  
|||||

QY 886 GCATCTGGTCAAGAAATTTATTTGGGTCACACAGATTTCTCTGCTAGTCCAGAAATCAT 945  
|||||

Db 278 GAACGTCATTAATGAGGTTTGAATGCTGTTAAATAAAGCTGAAACTGCTGCTAAATAGT 337  
|||||

QY 946 GAAGCCACAGTT-AAATTCGTTAGTTTATTAATCCACACCTTAACACGGTGATATCCA 1004  
|||||

Db 338 GAAGAACAAGTTCAATTTGGAGAAGATCTTTTGATACACCTCTCCACCATATGGAACCA 397  
|||||

QY 1005 TCATATGAATATTA 1018  
|||||

Db 398 GATCAATAATATTA 411  
|||||

RESULT 28  
BI503620

LOCUS  
DEFINITION  
mellifera cDNA clone BB170018A20E07 5', mRNA sequence.

ACCESSION  
BI503620  
VERSION  
BI503620.1  
KEYWORDS  
EST.  
SOURCE  
Apis mellifera (honeybee)

ORGANISM  
Apis mellifera  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Apoidea; Apidae; Apis.

REFERENCE  
1 (bases 1 to 657)  
Whitfield, C.W., Band, M.R., Bonaldo, M.F., Kumar, C.G., Liu, L., Pardinas, J., Robertson, H.M., Soares, B. and Robinson, G.E.

TITLE  
Annotated expressed sequence tags and cDNA microarrays for studies of brain and behavior in the honey bee

JOURNAL  
Genome Res. 12 (4), 555-566 (2002)

MEDLINE  
21929762

PUBMED  
11932240

COMMENT  
Contact: Gene E. Robinson  
Department of Entomology  
University of Illinois  
505 S. Goodwin Ave., Urbana, IL 61801, USA  
Tel: 217 265 0309  
Fax: 217 244 3499  
Email: generobi@life.uiuc.edu

This research was funded by the University of Illinois Critical Research Initiatives Fund and a Burroughs-Wellcome Trust Innovation Award in Functional Genomics to G.E. Robinson and an NSF Postdoctoral Fellowship in Bioinformatics to C.W. Whitfield.

PCR Primers

FORWARD: TAATACGACTCACTATAGG  
BACKWARD: ATTAACCCCTCACTAAG

Plate: BB170018A20 row: E column: 07

Seq primer: AGCGGATAACAATTCACACAGGA

High quality sequence stop: 657.

Location/Qualifiers

1..657

/organism="Apis mellifera"

/mol\_type="mRNA"

/strain="mixed strains of European bees, predominantly A.m. ligustica"

/db\_xref="taxon:7460"

/clone="BB170018A20E07"

/sex="female"

/tissue\_type="brain"

/dev\_stage="adult worker honey bee"

/lab\_host="DH10B"

/clone\_lib="Bee Brain Normalized/Subtracted Library, BB17"

/note="Organ: Brain; Vector: pT73-Pac; Site 1: EcoRI; Site 2: NotI; This BB17 cDNA library was generated by subtraction of the BB16 library with 4000 previously sequenced clones. The BB16 library was contributed by the Soares laboratory and it was constructed and normalized as described by Bonaldo, M.F., Lennon, G. and Soares, M.B. (1996). Genome Research 6(9): 791-806. RNA was prepared from dissected brains of adult worker bees of various ages and various behavioral groups."

BASE COUNT 255 a 91 c 114 g 197 t

ORIGIN

Query Match 4.6%; Score 51.6; DB 12; Length 657;  
Best Local Similarity 52.8%; Pred. No. 0.06;  
Matches 134; Conservative 0; Mismatches 119; Indels 1; Gaps 1;

QY 766 AAGTTCACACGACATGTCTACCTGCCACATTTAGAGTGTTCACATGACATTGAA 825  
|||||

Db 111 AAATTCATTGTTGCTCATACATCAFTATTATCAAGGCTCAAGTACTCTTAA 170  
Qy 826 AGCATTCCTGAAGAAATCAGAAAGATAATGTTAAATTTGATATCTTTGTTATCGT 885  
Db 171 GCAATTCCTTAAGAAATAGGCAAGAAACATTTACTGTTCAAAGACATGCGGATTAAT 230  
Qy 886 GCATCTGCTCAAGAAATTTCTATTTGGTGCAACAGATTTCTGCTAGTCCAGATATCAT 945  
Db 231 GAACGTCTATTGAGGTTTGACTGGTTTAAATAAAGCTGAACTGCTGCTAAATATGTT 290  
Qy 946 GAAGCCACAGTT-AAATTCGTTAGTTATATAATCCACCTTAAACGGTGATATCCA 1004  
Db 291 GAAGAACAAAGTTCAAATTTGGAGAAGATCTTTTGATACACCTCTCCACCTATGGAACCA 350  
Qy 1005 TCATATGAATATTA 1018  
Db 351 GATCATAAATATTA 364

RESULT 29  
BI506164  
LOCUS  
DEFINITION BB170014B20A07.5 Bee Brain Normalized/Subtracted Library, BB17 Apis mellifera cDNA clone BB170014B20A07 5', mRNA sequence.

ACCESSION BI506164  
VERSION BI506164.1 GI:15356538  
KEYWORDS EST.  
SOURCE Apis mellifera (honeybee)  
ORGANISM Apis mellifera  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Apoidea; Apidae; Apis.

REFERENCE 1 (bases 1 to 557)  
AUTHORS Whitfield,C.W., Band,M.R., Bonaldo,M.F., Kumar,C.G., Liu,L., Pardinas,J., Robertson,H.M., Soares,B. and Robinson,G.E.  
TITLE Annotated expressed sequence tags and cDNA microarrays for studies of brain and behavior in the honey bee  
JOURNAL Genome Res. 12 (4), 555-566 (2002)  
MEDLINE 21929762  
PUBMED 11932240

COMMENT  
Contact: Gene E. Robinson  
Department of Entomology  
University of Illinois  
505 S. Goodwin Ave., Urbana, IL 61801, USA  
Tel: 217 265 0309  
Fax: 217 244 3499  
Email: generobi@life.uiuc.edu  
This research was funded by the University of Illinois Critical Research Initiatives Fund and a Burroughs-Wellcome Trust Innovation Award in Functional Genomics to G.E. Robinson and an NSF Postdoctoral Fellowship in Bioinformatics to C.W. Whitfield.  
PCR Primers  
FORWARD: TAATACGACTCACTATAGG  
BACKWARD: ATTAACCTCTACTAAG  
Plate: BB170014B20 row: A column: 07  
Seq primer: AGCGATAACAATTTCACACGGA  
High quality sequence stop: 657.

FEATURES  
source  
1. .657  
/organism="Apis mellifera"  
/mol\_type="mRNA"  
/strain="mixed strains of European bees, predominantly A.m. ligustica"  
/db\_xref="taxon:7460"  
/clone="BB170014B20A07"  
/sex="female"  
/tissue\_type="brain"  
/dev\_stage="adult worker honey bee"  
/lab\_host="DH10B"  
/clone\_lib="Bee Brain Normalized/Subtracted Library, BB17"  
/note="Organ: Brain; Vector: pT73-Pac; Site: 1; Ecol; Site 2: Not1; This BB17 cDNA library was generated by subtraction of the BB16 library with 4000 previously

sequenced clones. The BB16 library was contributed by the Soares laboratory and it was constructed and normalized as described by Bonaldo, M.F., Lennon, G. and Soares, M.B. (1996), Genome Research 6(9): 791-806. RNA was prepared from dissected brains of adult worker bees of various ages and various behavioral groups. "

BASE COUNT 253 a 91 c 115 g 197 t 1 others  
ORIGIN  
Query Match 4.8%; Score 51.6; DB 12; Length 657;  
Best Local Similarity 52.8%; Pred. No. 0.06;  
Matches 134; Conservative 0; Mismatches 119; Indels 1; Gaps 1;

Qy 766 AAAGTCACAGCAGATAGTCTACCTGCCACATATTAGAGTGTTCCTCAATGACCATTTGAA 825  
Db 111 AAATTCATTTTGTATTTGCTCATATCATTTATTAACAGGCTCAAGTACTCTTAA 170  
Qy 826 AGCATTCCTGAAGAAATCAGAAAGATAATGTTAAATTTGATATCTTTGTTATCGT 885  
Db 171 GCAATTCCTTAAGAAATAGGCAAGAAACATTTACTGTTCAAAGACATGCGGATTAAT 230  
Qy 886 GCATCTGCTCAAGAAATTTCTATTTGGTGCAACAGATTTTCCTGCTAGTCCAGATATCAT 945  
Db 231 GAACGTCTATTGAGGTTTGACTGGTTTAAATAAAGCTCAAACTGCTGCTAAATATGTT 290  
Qy 946 GAAGCCACAGTT-AAATTCGTTAGTTATATAATCCACACCTTAAACGGTGATATCCA 1004  
Db 291 GAAGAACAAAGTTCAAATTTGGAGAAGATCTTTTGATACACCTCTCCACCTATGGAACCA 350  
Qy 1005 TCATATGAATATTA 1018  
Db 351 GATCATAAATATTA 364

RESULT 30

BI504612

LOCUS

DEFINITION BB170025B10C09.5 Bee Brain Normalized/Subtracted Library, BB17 Apis mellifera cDNA clone BB170025B10C09 5', mRNA sequence.

ACCESSION BI504612

VERSION BI504612.1 GI:15354986

KEYWORDS EST.

SOURCE Apis mellifera (honeybee)

ORGANISM Apis mellifera

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Apoidea; Apidae; Apis.

REFERENCE 1 (bases 1 to 669)

AUTHORS Whitfield,C.W., Band,M.R., Bonaldo,M.F., Kumar,C.G., Liu,L., Pardinas,J., Robertson,H.M., Soares,B. and Robinson,G.E.

TITLE Annotated expressed sequence tags and cDNA microarrays for studies of brain and behavior in the honey bee

JOURNAL Genome Res. 12 (4), 555-566 (2002)

MEDLINE 21929762

PUBMED 11932240

COMMENT  
Contact: Gene E. Robinson  
Department of Entomology  
University of Illinois  
505 S. Goodwin Ave., Urbana, IL 61801, USA  
Tel: 217 265 0309  
Fax: 217 244 3499  
Email: generobi@life.uiuc.edu  
This research was funded by the University of Illinois Critical Research Initiatives Fund and a Burroughs-Wellcome Trust Innovation Award in Functional Genomics to G.E. Robinson and an NSF Postdoctoral Fellowship in Bioinformatics to C.W. Whitfield.  
PCR Primers  
FORWARD: TAATACGACTCACTATAGG  
BACKWARD: ATTAACCTCTACTAAG  
Plate: BB170025B10 row: C column: 09  
Seq primer: AGCGATAACAATTTCACACGGA  
High quality sequence stop: 669.

FEATURES  
Location/Qualifiers

```
source
1. .669
/organism="Apis mellifera"
/mol_type="mRNA"
/strain="mixed strains of European bees, predominantly
A.m. linguistics"
/db_xref="taxon:7460"
/clones="BB170025B10C09"
/sex="female"
/tissue_type="brain"
/dev_stage="adult worker honey bee"
/lab_host="DH108"
/clone_lib="Bee Brain Normalized/Subtracted Library, BB17"
/notes="Organ: brain; Vector: pT73-Pac; Site 1: EcoRI;
Site 2: NotI; This BB17 cDNA library was generated by
subtraction of the BB16 library with 4000 previously
sequenced clones. The BB16 library was contributed by the
Soares laboratory and it was constructed and normalized
as described by Bonaldo, M.F., Lennon, G. and Soares,
M.B. (1996), Genome Research 6(9): 791-806. RNA was
prepared from dissected brains of adult worker bees of
various ages and various behavioral groups."
BASE COUNT 262 a 92 c 118 g 197 t
ORIGIN

Query Match 4.6%; Score 51.6; DB 12; Length 669;
Best Local Similarity 52.8%; Pred. No. 0.061;
Matches 134; Conservative 0; Mismatches 119; Indels 1; Gaps 1;

Qy 766 AAGTCACGCGAGATGCTACCTGCCACATTTATAGAGTGGTTCCTCAATGACCATTCGA 825
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 154 AAATTCACCTTTTGATATTGCTCATACATCATTTATATCAAGGGCTCAAGATACCTTTAAA 213
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 826 AGCATTCCTGAAGAAATCAGAAAAGATAATGTTAAATTTGATATTACTTTGTTGTCATCGT 885
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 214 GCATTTCTTAAGAAATAGACAGAAAACATTTACTGTTCAAAGACATGCGGATTAAT 273
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 886 GCATCTGGTCAAGAAATCTATTGGGTGCAACAAGATTTCTCTGTAGTCCAGAAATATCAT 945
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 274 GAAGTCATTATGCGAGGTTTGACTGGTTTAAATAAAGCTGAAGCTGCTGCTAAATATGGT 333
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 946 GAAGCCACAGTT-AAATTCGTTAAGTTTATAAATCCACACCTTAACACGCGTATATCCA 1004
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 334 GAAGAACCAAGTCAAAATTTGGAGAAGATCTTTTGATACACCTCTCCACCTATGGAACCA 393
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1005 TCATATGAATATTA 1018
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 394 GATCATAATATTA 407
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Search completed: January 10, 2004, 08:39:36  
Job time : 2789 secs

**THIS PAGE BLANK (USPTO)**